

The QED Group, LLC

IHSS Monitoring and Evaluation Training – DRAFT TRAINING MODULES

**Subcontract No. 11155, TA #2
Work Plan Reference No. 1.6.2**

February 15, 2004

Prepared for:

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Submitted By:

The QED Group, LLC
1250 Eye Street, NW
Washington, DC 20005
(202) 898-1910

PROPOSED AGENDA FOR 2-WEEK M&E WORKSHOP

Day 1 (AM):	Course Overview
Day 1 (PM):	The Monitoring and Evaluation Process
Day 2 (AM):	Planning the M&E Process – Identifying the Questions
Day 2 (PM):	Developing a Project Intervention
Day 3 (AM):	Setting Performance Targets
Day 3 (PM):	Measuring Performance Results
Day 4 (AM):	Measuring Performance Results (Cont.)
Day 4 (PM):	Selecting the M&E Design Approach
Day 5 (AM/PM):	Impact Evaluation Designs
Day 6 (AM/PM):	Collecting Data on M&E Indicators
Day 7 (AM):	Basic Sampling Issues
Day 7 (PM):	Analyzing Performance Data
Day 8 (AM):	M&E Qualitative Data
Day 8 (PM):	M&E Quantitative Data
Day 9 (AM):	Cost Effectiveness Analysis
Day 9 (PM):	Using the M&E Results
Day 10 (AM/PM):	Planning the Logistics of Monitoring and Evaluation
Day 11 (AM):	Example M&E Plan
Day 11 (PM):	Preparing Group M&E Plans
Day 12 (AM):	Presentation of Group M&E Plans and Discussion
Day 12 (PM):	Training Course Wrap-Up Session

Day 1 (AM): COURSE OVERVIEW

1. Introduce trainers and provide brief summary of their M&E experience
2. Purpose of the training course

Iraq health officials need to develop the capacity to monitor and evaluate the projects they are responsible for implementing at the governorate level. Increasing this capacity will improve health project performance and the performance of the health system as a whole. The training course is designed to meet this need by discussing the key components of a health monitoring and evaluation (M&E) system and having the participants develop a M&E plan for potential application to their health projects.

3. Present the course agenda and discuss the participative requirements of the course. How are the trainees expected to participate in various exercises?

The course combines lectures and course participant exercises to present the training material and check on its comprehension as the course progresses. The lectures will contain material from M&E literature and trainer experience. The exercises will provide opportunities for the participants to demonstrate their mastery of the lecture material and receive feed back from the trainers.

4. Discuss the course materials and how they will be used.
5. Preview of the day 1 lecture
 - The need to assess the readiness of your health agency for M&E as a first step in developing a M&E planning and implementation.
6. Assessing the readiness of a public health agency for developing a M&E capacity.
 - How are projects monitored and evaluated now?
 - Is M&E data routinely collected?
 - Is this information used for M&E? How is it used?
 - Is there a demand for more M&E information?
 - Is there a demand for better and more useful M&E information?
 - Does the staff have the skills and experience to do quality M&E?
 - Are there any major obstacles to doing M&E in your health organization?

Interactive session: What do participants see as the purpose and usefulness of M&E for the health projects? How would it help them in program/project decision making? Is there organization ready for using M&E planning?

- What is needed to overcome these obstacles?

DAY 2 (PM): THE MONITORING AND EVALUATION PROCESS

1. Preview of the day 2 lecture

- The necessity of M&E capacity as a basis for managing public health agencies to produce intended results.
- The concept of results-based management

2. Results-Based Management (RBM)

RBM is a means to improve management effectiveness and accountability by involving key stakeholders in defining realistic expected results, assessing risk, monitoring progress toward the achievement of expected results, integrating lessons learned into management decisions and reporting on performance.

The focus is on results –

Were all of the key components of the health project fully implemented?

Did these activities produce their intended results?

3. What is the value of M&E to results-based management?

- Accountability
 - Is the project achieving its stated goals and objectives? (e.g., improved health outcomes)
 - In what ways is the project performing best? (e.g., reaching selected population sub-groups)
 - In what ways is the project not meeting its goals and objectives? (e.g., not fully implementing a certain part of the program)
 - Are there any barriers to the project meeting its goals and objectives? Can they be overcome?
- Project improvement
 - How can the performance of the project be improved?
 - What specific actions can be taken to improve project performance?
 - What is needed to take these actions?
 - Are there any limits to improving project importance?
 - What can be done to deal with these limits?

- Efficient use of scarce resources
 - Is the project using its resources according to its implementation plan?
 - Is the project staying on-budget?
 - What have been the major obstacles to the proper use of these resources?
 - Does the project have sufficient resources to meet its goals and objectives?
 - Are other resources needed?
 - How would additional resources be used to improve project performance?

3. Overview of essential M&E components

- Implementation analysis
 - Focus on the implementation of the main project activities
 - Documentation of project resource utilization
 - Identification of break downs of and/or delays in project implementation
 - Assessment of the effects of these break downs/delays on project implementation
 - Unforeseen obstacles to project implementation and how to overcome them
- Impact analysis
 - What specific changes can be traced to project activities?
 - Do these changes match project goals and objectives?
 - Are there other plausible explanations for these observed changes?
 - Factors other than project activities
 - To what extent overall has the project produced its intended impacts?
 - In what specific areas or ways has the project not produced its intended impacts?

- Managing the M&E process
 1. Project M&E staffing
 - a. Who is going to do the M&E activity?
 - b. What skills are needed?
 - c. What will be their responsibilities
 - d. Who will manage the process?
 2. Project budgeting for M&E
 3. Scheduling the M&E activities
 4. Timely decision making for M&E management
- Criteria for credible and useful M&E plan
 1. Clear statement of project goals and objectives
 2. Logic model for the project intervention
 3. Useful indicators of project implementation and impact
 4. Reliable measurement of indicators
 5. Credible basis for M&E comparisons
 6. Sound analysis of M&E data
 7. Timely and useful reporting of M&E results
 8. Use of M&E results in project decision making

Theory-Based M&E

The premise of theory-based M&E is that programs and projects are based on explicit or implicit theory about how and why a program will work. The M&E would then be based on assessing each theory and assumptions about a program during implementation rather than at a midpoint or after the project has been completed. In designing the M&E, the underlying theory is presented as many microsteps, with the methods then constructed for data collection and analysis to track the unfolding of assumptions. If events do not work out as expected, the M&E can say with a certain confidence where, why, and how the breakdown occurred.

The approach puts emphasis on the responses of people to program activities. Theories direct the evaluator's attention to likely types of near-term and longer-term effects. Among the advantages are, first, that the M&E provides early indications of program effectiveness during project implementation. If there are breakdowns during implementation, it is possible to fix them along the way. Second, the approach helps to explain how and why effects occurred. If events work out as expected, the M&E can say with a certain confidence how the effects were generated.

DAY 2 (AM): PLANNING THE M&E PROCESS: IDENTIFYING THE QUESTIONS

The first step in the M&E planning process is deciding which questions to ask. Good questions are clear, focused, and relevant. The questions make sense to others. Deciding on which questions to ask is not easy since there are many possible questions that can be asked about an intervention. Asking a slightly different question can take you down a different pathway in terms of the data needed and the methods for collecting it. For example, say you have been asked by a donor to evaluate an intervention that educated village families about child health. If you ask a question about changes in viral infection rates, you might want to bring in some medical experts to determine the number of people infected before and after the intervention. If you ask a question about the content of pamphlets and information sessions, you might have experts check it for accuracy. If you want to know how well the information was communicated, you would want to talk with the villagers to gauge their understanding of the message, and their sense of its appropriateness/effectiveness in their culture.

When deciding which questions to ask, you want to consider several factors:

1. What is the main need that the intervention is attempting to address?
2. Who are the likely users of your M&E study and what are they most interested in? This is sometimes called a stakeholder analysis. Their interests may vary and you may need to ask several different questions to accommodate the different issues and concerns.
3. What is the relative importance each of the possible questions?
4. What are the relative ease, resource requirements and data availability for answering the possible questions?

When deciding on the M&E questions, you need to ask yourself:

1. What does the donor need to know?
2. What do the other stakeholders need to know?
3. Where is the intervention in terms of its life cycle?
4. How will the results be used?
5. Which questions are most important?

Possible Sources to Identify the Questions

M&E questions may vary depending upon the focus: indicators of performance, activities, implementation, results, impacts, relevance, efficiency and effectiveness, impact and/or sustainability. If you are conducting a goal-based M&E, it will be helpful to review documents to see if the goals and objectives of the intervention were specified. You will also want to see if other studies have been done: what questions did they ask and what did they learn? You will want to determine whether specific M&E questions were part of the funding agreements. You will need to find out what the various stakeholders think are important questions. Experts in the field might also have useful insights.

Remain open to a wide variety of possible questions and resist locking into a question before you have read the necessary material and have spoken with stakeholders and experts in the field. Keep a running list of questions with possible ideas about how they might be answered.

Stakeholder Analysis

Since deciding on the M&E questions is the first step in planning the M&E design, stakeholders should be involved. Stakeholders are all those who people or representatives of organizations that have a “stake” in the intervention. Typically they are those who affected by an intervention either during its lifetime or in subsequent years. It is important to include those who would typically not be asked to participate.

Stakeholders can include:

1. Participants: those people who participate or have participated in the intervention.
2. Beneficiaries: those people who directly and currently benefit.
3. Indirect Beneficiaries: those people who are not recipients of the intervention but who benefit from others who are beneficiaries. For example, employers benefit from educational programs since they are able to hire better-trained people.
4. Other Impacted: those people who did not participate in the program but who were impacted by it in some way, either positively or negatively.
5. Donors.
6. Government officials, elected officials, government employees with a relevant interest, such as planners, public health nurses, etc.
7. Program directors, staff, board members, managers, volunteers.
8. Policy-makers.

9. Community and interest groups, including those that might have a different agenda from the program officials.

While it may be somewhat unwieldy, involvement of stakeholders in this first step is likely to:

- Generate better questions
- Generate support for the M&E
- Increase access to whatever information is available
- Enhance the acceptance of the final report and recommendations.

By engaging the stakeholders early on, everyone will have a better understanding of the intervention and the challenges it faces in implementation. In addition, the M&E team will be better informed about what information is needed, when, and by whom.

The extent to which stakeholders are actively involved in the design and implementation of the M&E depends on several factors. For example, stakeholders may not be able to afford to take time away from their regular duties, or there may be political reasons why the M&E needs to be seen as independent.

Stakeholders: Diverse Perspectives

Stakeholders approach the intervention from different perspectives. This is a good thing. It helps to understand that the initial discussions may reflect those perspectives. A donor may be concerned that the money is spent appropriately and that the intervention is effective. A program manager may be concerned that the intervention is well managed and is successful. Program participants may want to get more and/or better services. Policy-makers may be most concerned with whether the intervention is having an impact. Others in the community may want to expand the intervention, while others may want to limit what they perceive to be some of the negative consequences of the intervention.

Disagreement is a normal part of the process of people working together. People who feel passionately often have somewhat different visions of how the world is and should be. As a facilitator, it is important for the evaluator to help the group set ground rules about disagreement that make sense within the cultural context. But it is essential that disagreement about issues and ideas be brought into the open, discussed and resolved in a way that everyone feels is fair.

Involving Stakeholders

The first challenge will be to identify the stakeholders. This can be done by looking at documents about the intervention and talking with program staff, local officials and program participants. While stakeholders can be interviewed initially, involvement of stakeholders is a group process.

Field visits are essential. Either periodic stakeholder meetings can be held or a more formal structure can

be established, such as an advisory or steering committee structure. Tasks can be assigned to individuals or to smaller sub-committees if necessary.

In a fully participatory M&E the evaluator serves as a facilitator, enabling all participants to share their views and ideas. A facilitator is responsible for managing the process:

1. To set an agenda
2. To help the group stick to the agenda (topics and times schedule)
3. To ensure that all views are heard
4. To oversee a process for decisions making (a consensus or a voting process).

While the evaluator has the technical expertise about how to do the M&E, the stakeholders in a participatory M&E are usually allowed to shape the focus of the M&E. In a sense, the analyst serves as staff for these kinds of M&Es. This means that the evaluator, who listens and facilitates the discussion about M&E focus, can summarize, prepare written notes, and provide stakeholders with options about ways the M&E can be approached.

Interactive session: Who are the key stakeholders for your program or project? Make a list of them. What are their priority information needs? For what decisions do they need M&E information? How would they use the

DAY 2 (PM): DEVELOPING A PROJECT INTERVENTION

Developing and using a logic model

You need to start your M&E process with a clear understanding of your project's goals and objectives. Next you need to think about the activities your project does, and your beliefs about how those activities will result eventually in reaching your project's goals. The following steps can guide the process:

- [**Step 1. Make a list of project goals.** Because it is often difficult to measure and document long-term goals such as reducing violence against women in your community, you want to be sure to start with short-term goals, then include some medium-term and long-term goals also. Short-term goals could include, for example (1) increasing the number of children getting vaccinations; (2) increasing the number of mothers taking part in nutrition education classes; or (3) getting more women in underserved groups to approach your service.

Many projects have found that the process of discussing and listing goals is useful in clarifying differences in opinions among project staff and participating agencies on goals.

- [**Step 2. Make a list of project services and other activities.** Activities can include project planning steps such as convening meetings or reviewing training curriculum materials, as well as the delivery of services or products such as training people in the in-patient process.
- [**Step 3. Make a list of all the background characteristics** (factors) of the people involved that you think may influence the relationship between activities and goals. For example, training may have different effects depending on the rank or years of experience of nurses trained; the impact of patient services may depend on the person's age, number of children, or general health.
- [**Step 4. Make a list of all the events or factors happening during or after your project activities** that could influence how or whether your project accomplishes its goals. In the case of a program designed to combat domestic abuse, for example, these could include whether a victim has continued contact with an abuser, or whether law enforcement officers are placed in positions that use the information given during a training program. Exhibit 1 displays how this type of information could be organized for a domestic abuse prevention program.

EXHIBIT 1
LOGIC MODEL FOR EVALUATION OF SHELTER-BASED SERVICES
FOR DOMESTIC VIOLENCE VICTIMS (SAMPLE EXERCISE)

Column A	Column B	Column C	Column D	
Background Factors	Program Services and Activities	External Services/Factors	Goals (Outcomes)	
			D1	D2
			Immediate	Long-Term
Children	Shelter/Housing	Policy Response	A Safety Plan	Reductions in
Language	Temporary			-threats/stalking
	Transitional			-emotional/psychological abuse
History of Abuse	Counseling	Family/	Immediate Safety	-physical abuse
	Individual	Friends/Social		-injury
	Group	Support		
Employment/ Education/ Income	Emergency Assistance	Availability of Needed Services	Linkages to Services as Needed	Increases in:

Column A	Column B	Column C	Column D	
Background Factors	Program Services and Activities	External Services/Factors	Goals (Outcomes)	
	-cash -food -clothing	Court Response	D1	D2
			Immediate	Long-Term
			-housing -health care -job/education -legal assistance	-perceived safety -empowerment -mental health
Pending Legal Action	Legal Advocacy Court accompaniment Help with protection orders Referrals		Help with Children's Needs Counseling Custody/visitation Health care Day care	
			Increased Legal Protection	

EXHIBIT 2
LOGIC MODEL FOR YOUR PROGRAM

Column A	Column B	Column C	Column D	
Background Factors	Program Services and Activities	External Services/Factors	Goals (Outcomes)	
			D1	D2
			Immediate	Long-Term

The diagrams can be used to plan your evaluation as follows:

- In impact evaluations, the diagram is used to spell out how, and for whom, certain project activities are expected to attain specific goals. This process involves developing a set of hypotheses to be tested in the analysis. Simple hypotheses will usually take the form of "A service from Column B leads to an outcome in Column D." More complex hypotheses might take the form of "Column B leads to Column D, but happens faster/more if conditions in Column C are right." Sample hypotheses for the logic model presented in (Exhibit 2) might be:
 - Participating in group and individual counseling helps women become more empowered.
 - Receipt of legal advocacy increases legal protection and immediate safety.
 - Legal advocacy increases legal protection and immediate safety, but will do it better if the police and court responses are supportive, and will have trouble doing it if these responses are obstructive.
- In process evaluation, the diagram is used to identify expectations about how the project should work—an "ideal type" — which can then be used to assess the deviations in practice, why these deviations have occurred, and how the deviations may affect achievement of goals either positively or negatively. This vision of the program assists project managers and evaluators to identify differences from expectations (including positive and negative unintended consequences), consider possible ways to fine-tune project operations so they look more like the planned approach, or go back to the drawing board to consider changes to the original approach.
- In performance monitoring, the diagram is used to focus on which kinds of indicators (statistics) are appropriate for specific target populations, communities, or time periods. The process of making the list of strategies and outcomes often helps projects identify gaps in planning or differences among staff and partner agencies in assumptions about how the project will operate and the various responsibilities of participants.

The logic model can also be the basis for preparing a list of variables (indicators) needed to measure the concepts identified in each box. To do this, make a separate page for each box in your own logic model. Take the first page and list the variables needed for the first box. Using the example in Exhibit 1, the first box is "Children." Basic variables might be: yes/no; number of children; and, for each child, age, gender, where child lives (with client, where else). Additional variables that might be relevant depending on what types of services your program offers, might be whether child were also abused, how they were abused, what school they are attending, whether they have attendance problems, etc.

Once you have listed variables, then list next to each variable the source(s) where you will (may) be able to get the information, such as court records, program records, interview with the victim, the program staff, etc. Whenever possible, we recommend getting information on the same variable from multiple sources, to verify that the information is correct (e.g., number of children) or to have several "reads" on the same variable if it is a hard one to measure. For example, you could measure reductions in victimization with information from both the victim and police records—each would tell you something different about the outcome.

Repeat these two steps (listing variables and listing data sources) for each box in your own logic model (Exhibit 2).

Once you know what variables you want to measure and where the information will come from, you can begin to develop data collection plans. These can take a while. You need to find out what records are maintained by agencies, decide what information is needed to construct variables, get an understanding of the limits and meaning of the records, arrange permission to get the data if you don't already have it all in your files, and develop a protocol specifying exactly what data elements will be collected, by whom, and how. You need to decide when and how program participants can be contacted to provide information, what questions should be asked, and who will collect the information. Similarly, you will want to make lists of persons on the program staff and working with other agencies who need to be interviewed and what questions are appropriate for each. You may want to add to or change the forms your own agency uses to be sure that you are collecting the necessary information for each client.

Interactive Session:
participant groups
develop health project
logic model. They go as
far as they can in filling in
the model. Participants
and trainers critique logic
models.

DAY 3 (AM): SETTING PERFORMANCE TARGETS

Performance Targets

Performance targets are quantified objectives, set by the health project management, to be attained at a future date. They express the aims of the process, at any level, and provide the basis for identifying problems and moving towards solutions as early as possible. Setting targets, objectives or goals is essential, otherwise there is no basis for choosing what to measure, how to assess the information obtained, or what action to take. Defining a target answers the question 'what are we aiming for?'

Targets can take different forms, depending on the process being assessed. For complex processes, there may be several targets, at different stages, and it may be that meeting targets at lower levels is a vital, enabling step towards meeting higher-level targets.

Benchmarking

One increasingly popular way of setting targets is to look at what is being done by someone else -- another health department or group of health departments -- that has a reputation for high performance in the provision of health services.

Some examples are simple. How much time should be spent at patient intake to get vital health information? Similarly, hospital operating units may seek such benchmarks in a particular program area by examining the best experiences of others --e.g., other hospitals handling similar patient loads -- that have achieved a high level of performance.

Targets may be set to reflect this "best in the medical practice" experience, provided of course that consideration is given to the comparability of local conditions, resource availabilities, and other factors likely to influence the performance levels which can be achieved.

Some kinds of targets, with examples of each, are:

- fulfilling a binary (yes/no) measure, often within a set timescale (Did project participants get the health services? Did their health score improve?)
- achieving a set level of input (health costs reduced by a certain percentage each year)
- achieving a set level of output (number of patients showing health improvement)
- achieving a set level of quality (number of health service complaints below a set level)
- realizing outcomes (e.g., improved health, disease reduction)
- Binary and input targets may be important but have no regard to quality. Make sure quality targets are included. (e.g., level of patient satisfaction with service received).

[**Concept of % target achievement**

A simple way to assess target achievement is examines the % of achievement. This can be done by using the following formula:

$$\% \text{ Target Achievement} = (\text{Actual} - \text{Baseline}) / \text{Target}$$

where:

Actual = Performance score achieved to date
Baseline = Performance score at project start-up
Target = Performance score the project hopes to achieve

Example: Weight loss

Actual = weight as measured after six months project participation (75 Kilos)

Baseline = weight as measured at the beginning of the project (85 Kilos)

Target = desired weight loss at end of project participation (15 Kilos)

$$\begin{aligned} \% \text{ Target Achievement} &= (75-85)/70 \\ &= (-10)/15 \\ &= -66\% \text{ weight loss} \\ &= 66\% \text{ target achievement} \end{aligned}$$

[**Criteria for useful performance targets**

1. Good performance targets will be SMART: Specific, Measurable, Achievable, Relevant and Timed.
2. specific: clear, unambiguous and easy to understand by those who are required to achieve them
3. measurable: there is no point setting a target for which success cannot be gauged by referring to a specific measure or measures
4. achievable: expressing specific aims that staff feel can realistically be achieved, with some effort: 'out of reach, but not out of sight'
5. relevant to those who will be required to meet them; they must have enough control over their work to be able to meet their targets, or their motivation will suffer
6. timed: there should be a set timescale for achieving a target; open-ended targets may not encourage focused effort on improving performance.

If targets are hard to quantify, the definitions of the terms used to describe them are crucial and should be agreed between all concerned. Terms such as 'satisfaction' and 'milestones' are open to different interpretations and may have to be precisely defined for your situation.

Exercise: Participants develop performance targets for particular health service. For example, targets for maternal and child health clinics. Trainers critique the targets.

DAY 3 (PM): MEASURING PERFORMANCE RESULTS

Following the identification of performance targets, the next task is to select performance indicators that can be used to measure target achievement. It is important that the stakeholders agree a priori on the indicators that will be used to measure program/project performance. Performance indicators are qualitative or quantitative measures of resource use, extent of program coverage and health results achieved

that are used to monitor program/project performance. Quantitative indicators are statistical measures such as number, frequency, percentile, ratios, variance, etc. Qualitative indicators are judgment and perception measures of congruence with established standards, the presence or absence of specific conditions, the extent and quality of participation, or the level of health service recipient satisfaction, etc.

It is a popular myth that information collected on quantitative indicators is inherently more objective than that collected on qualitative indicators. Both can be either more or less objective or subjective depending on whether or not the principles of sound research have been rigorously applied in the data collection and analysis process.

Considerations/Limitations

Selection of performance indicators entails several practical issues

The set of performance indicators should be simple, limited to a few key indicators of priority outcomes. It is recommended to collect the “minimal essential data” on vital few indicators. Too many indicators burden the data collection and analysis and make it less likely that managers will understand and use reported information. At the same time, the set of indicators should be constructed to reflect the informational needs of stakeholders at all levels—community members, agency directors, and national funders. Most importantly, the performance indicators should reflect key activities defined as central to the project in the logic model.

Regular measurement, at least quarterly, is important so that the system provides the information in time to make shifts in project operations and to capture changes over time. However, pressures for timely reporting should not be allowed to sacrifice data quality. For performance monitoring to take place in a reliable and timely way, the evaluation should include adequate support and plans for training and technical assistance for data collection. Routine quality control procedures should be established to check on data entry accuracy and missing information. At the point of analysis, procedures for verifying trends should be in place, particularly if the results are unexpected.

The costs of performance monitoring are modest relative to impact evaluations, but still vary widely depending on the data used. Most performance indicator data come from records maintained by service providers. The added expense involves regularly collecting and analyzing these records, as well as preparing and disseminating reports to those concerned. This is typically a part-time work assignment for a supervisor within the health agency. The expense will be greater if client satisfaction surveys are used to measure outcomes. An outside survey organization may be required for a large-scale survey of past

clients; alternatively, a self-administered exit questionnaire can be given to clients at the end of services. In either case, the assistance of professional researchers is needed in preparing data sets, analyses, and reports.

Exercise: participants begin identifying performance indicators that fit their performance targets. They should select measures of inputs, outputs, outcomes and impacts. Trainers will be available as resource persons as they begin the

DAY 4 (AM): MEASURING PERFORMANCE RESULTS (CONT.)

There are six criteria that should be used when selecting performance indicators. Each one is presented below along with an illustrative question in guise of an explanation.

1. Validity - Does it measure the result?
2. Reliability - Is it a consistent measure over time?
3. Sensitivity - When the result changes will it be sensitive to those changes?
4. Simplicity - Will it be easy to collect and analyze the information?
5. Utility - Will the information be useful for decision-making and learning?
6. Affordability - Can the program/project afford to collect the information?

Performance indicators should be identified across the entire spectrum of the performance framework, from resource inputs through to project impacts or results. A minimalist approach to measuring resources would be advised by tracking financial expenditures by program/project component. Gender, age, profession, income, geographic location (rural/urban) and other indicators are generally useful when measuring the extent of program coverage. The choice of performance indicators to measure the achievement of results, especially at the output and outcome levels, will depend wholly on the nature of the result, how it is articulated and the implementation context including cost, level of effort, the size and complexity of the program/project.

At the outcomes level, the information collected on performance indicators could be analyzed and used in management decision-making to keep a program/project on track toward the achievement of its purpose. Information collected on the same indicators would also constitute evidence regarding program/project success, or failure at termination. It is suggested that at least three indicators per expected result at the outcomes level should be used: at least one quantitative, one qualitative and one of your choice. In many cases, a total of two indicators at the output level would be sufficient. For each quantitative indicator, it is important to specify the unit of analysis or calculation, existing baseline information and useful benchmarks for comparison. Benchmarks should also be specified for each qualitative indicator as well as expected perceptions or judgment of progress by stakeholders and a detailed description of expected conditions or situation to be observed.

Program/project stakeholders should begin the process of identifying and selecting performance indicators by preparing a comprehensive list. The next step is to decide how many are needed and apply the selection criteria above to the list. Those that don't meet these criteria should be discarded. The best performance indicators from those remaining should be used and the rest kept in a reserve pool. Some performance indicators may, after some use, prove not to meet the above criteria and must then be replaced from the reserve pool.

It may instructive to discuss of variety of health performance Indicators. This will provide practical examples that may be useful in thinking about your own indicators. These examples were taken from the National

1. Access to a GP

Percentage of patients able to be offered a routine appointment to see a GP within two working days

Rationale

The NHS Plan states that by 2004, patients will be able to see a primary care professional within 24 hours and a GP within 48 hours. In future, all practices will be required to guarantee this level of access for their patients, whether by providing the service themselves, or by entering into an arrangement with another practice, or by the introduction of

further NHS walk-in centers or other walk-in services.

Data source Data Period

Primary Care Access Survey To be finalized

2. Access to a primary care professional (PCP)

Percentage of patients able to be offered a routine appointment to see a primary care professional within one working day

Rationale

The NHS Plan states that by 2004, patients will be able to see a primary care professional within 24 hours, and a GP within 48 hours. In future, all practices will be required to guarantee this level of access for their patients, either by providing the service themselves, or by entering into an arrangement with another practice, or by the introduction of further NHS walk-in centers or other walk-in services. If the nurse is not available, then an appointment with the GP counts.

Data source Data Period

Primary Care Access Survey To be finalized/Commission for Health Improvement

3. Drug misuse: treatment

Increase in drug misusers accessing treatment

Rationale

This indicator is directly related to a Department of Health Target of Improving drug treatment, recognized by Ministers and the Prime Minister as the lynchpin of the national drugs strategy. This is based on evidence showing the dramatic effects that access to effective drug treatment can have for the individual and in reducing crime. These benefits include substantial financial savings within both the criminal justice system via reduced offending and the NHS through reduction in blood-borne diseases amongst drug misusers, and also the other associated health costs that chaotic drug misuser will account for. Reducing Drug Misuse is one of the key priorities in the Planning and Priorities Framework (PPF) for 2003-2006. PCTs have a key role, with other partners, in Commissioning integrated treatment and prevention programs and in implementing national guidance. The PPF includes the PSA target to increase the participation of problem drug users in treatment programs by 55% by 2004 and 100% by 2008 (from a 1998 baseline). PCTs have been asked to plan to achieve this in their Local Delivery Plans.

Data source Data Period

LDPR special collection/National Drug Treatment Monitoring System
To be finalized 2002/03 and 2003/04

4. Financial management

Achievement of the financial position shown in the 2003/04 Plan, submitted to the Department of Health, without the need of unplanned financial support

Rationale

Financial stability is both a key objective and a minimum standard for NHS bodies. It provides the essential platform on which to manage and develop patient services in line with the targets/objectives set out in the NHS Plan.

Data source Data Period

Month 12 Forecast Financial Returns Financial Year 2003/04/Commission for Health Improvement

5. Four-week smoking quitters

Number of smokers who quit at four-week follow-up with the NHS smoking cessation services (performance against plan)

Rationale

Smoking is the single greatest cause of preventable illness and premature death in the UK. Seven out of ten adult smokers say they would like to give up, but due to the addictive nature of nicotine, most smokers find it hard to quit. For smokers who give up, the chances of getting a serious or fatal disease are greatly reduced. This indicator is therefore crucial to securing improvements in public health. The Priorities and Planning Framework for the NHS for 2003-2006 sets a target of 800,000 four-week smoking quitters and makes clear the role of the services in contributing to reducing cancer mortality and in tackling health inequalities, through delivery of this target.

Data source Data Period

Smoking Cessation Return Financial Year 2003/04

6. Improving Working Lives

Continued implementation of the Improving Working Lives standard

Rationale

The NHS Plan states that improving the working lives of staff contributes directly to better patient care through improved recruitment and retention and because patients want to be treated by well-motivated, fairly-rewarded staff. The NHS Plan set out a commitment by NHS employers to implement more flexible, supportive and family friendly working arrangements that improve diversity, tackle discrimination and harassment, and develop the skills of all its staff to improve patient services.

Data source Data Period

HRD-EMP IWL Administration System Financial Year 2003/04/Commission for Health Improvement

7. Outpatients waiting longer than the standard

Measurement of the breaches of the 21-week target for first outpatient appointment throughout the year and the measurement of the breaches of the 17-week target for first outpatient appointment as at 31 March 2004

Rationale

Public consultation prior to the production of the NHS Plan indicated that the public wanted to see reduced waiting times in the NHS. The NHS Plan set the target to have no patients waiting longer than 21 weeks for an outpatient appointment, following a GP referral, by March 2003, and no patients waiting longer than 17 weeks by March 2004. Urgent cases will continue to be treated in accordance with clinical need.

Data source Data Period

QM08 Quarterly Waiting Times Returns for PCTs/Financial Year 2003/04

8. Patients waiting longer than the standard for elective admission

Measurement of the breaches of the 12-month target for an elective inpatient or day case admission throughout the year and the measurement of the breaches of the 9-month target for an elective inpatient or day case admission as at 31 March 2004

Rationale

Public consultation prior to the production of the NHS Plan indicated that the public wanted to see reduced waiting times in the NHS. The NHS Plan set the target to have no patients waiting longer than 12 months for an inpatient appointment by March 2003, and no patients waiting longer than 9 months by March 2004. Urgent cases will continue to be treated in accordance with clinical need.

Data source Data Period

QF01 Quarterly Waiting Times & Monthly Waiting List/Times returns
To be finalized Commission for Health Improvement

9. Total time in A&E: less than 4 hours

Percentage of patients waiting less than four hours in A&E from arrival to admission, transfer or discharge

Rationale

The NHS Plan target requires that no patient spends more than 4 hours in any type of A&E from arrival to admission, transfer or discharge by December 2004. Progress to the target is measured as a key indicator. From March 2003, 90% of patients should spend no more than 4 hours maximum in A&E.

Data source Data Period

QMAE quarterly return (A&E)/Sit Reps/To be finalized /Commission for Health Improvement

Balanced Scorecard Indicators

10. CAMHS

Child and adolescent mental health services (CAMHS): (i) audited needs assessment for the population (ii) increase in investment in CAMHS against target expected

Rationale

There is a Public Service Agreement target to increase CAMHS by at least 10% each year across the service according to agreed local priorities (demonstrated by increased staffing, patient contacts and/or investment).

Data source Data Period

CAMHS service mapping To be finalized

11. Cervical screening

Percentage of women aged 25-64 screened for cervical abnormalities

Rationale

The cervical screening programme screens nearly 4 million women each year in England, and experts estimate that the programme prevents up to 3,900 cases of cervical cancer each year. The programme between 1988 and 1997 saved over 8,000 lives. The incidence of cervical cancer has fallen by 43 percent between 1988 and 1997.

Data source Data Period

KC53 Cervical Screening To be finalized

12. CHD register

Clinical audit data that describe the provision of appropriate lifestyle advice and systematic treatment regimes

Rationale

The Priorities and Planning Framework states that primary care should update practice-based-registers so that patients with CHD and diabetes continue to receive appropriate advice and treatment in line with National Service Framework standards.

Data source Data Period

LDPR special collection To be finalized/Commission for Health Improvement

13. Child protection

Compliance to recommended child protection systems and procedures

Rationale

Lord Laming's inquiry into the death of Victoria Climbié stated that within the NHS, the organizational systems were not in place to ensure continuity of care or adequate consultant supervision. Child protection is a highly emotive subject, evoking strong feelings in all of us. In order to provide the most effective support and help for vulnerable children, young people and their families, all staff managing services and involved in working directly with children and young people need to be able to acknowledge their feelings

and examine their values and beliefs.

Data source Data Period

CHI self-assessment audit tool As at 31st March 2004

14. Community equipment

Percentage of items of equipment and adaptations delivered within seven days

Rationale

Equipment plays a vital role in enabling disabled people of all ages, including children, to maintain their health and independence, and to prevent inappropriate hospital admissions. Improving the service has links with both the NSF for Older People and with Valuing People, the strategy for services for people with learning disabilities.

Data source Data Period

LDPR special collection To be finalized

15. Death rates from cancer, ages under 75 (change in rate)

Percentage change in mortality rates from all malignant neoplasms in people aged under 75 per 100,000 population (age and sex standardized)

Rationale

Cancer is one of the most common causes of death in this country and is responsible for one out of every four deaths (almost 124,000 each year). An even greater percentage of deaths occur at younger ages, more than one in three deaths under the age of 75. There is much that can be done to reduce the death rate from cancers. Cancer was highlighted as a national priority area in the Saving Lives: Our Healthier Nation White Paper, the NHS Plan and the Department's Public Service Agreement. The target is to reduce the death rate in people aged under 75 years by at least one-fifth by 2010.

Data source Data Period

ONS Calendar Years 2002 and 2003/Commission for Health Improvement

16. Death rates from circulatory diseases, ages under 75 (change in rate)

Percentage change in mortality rate from all circulatory diseases in persons aged under 75 per 100,000 population (age and sex standardized)

Rationale

Circulatory diseases are a major cause of early death, accounting for 39,000 deaths in men (a third of all deaths) and 20,000 deaths in women (a quarter of all deaths) aged under 75 years annually. Heart disease and stroke can often be prevented. Circulatory diseases were highlighted as a national priority area in the Saving Lives: Our Healthier Nation white paper and the Department's Public Service Agreement. The target is to reduce the death rate in people aged under 75 years by at least 20% by 2010.

Data source Data Period

ONS Calendar Years 2002 and 2003

17. Delayed transfers of care

Percentage of patients whose transfer of care from hospital was delayed

Rationale

This indicator measures the impact of community-based care in facilitating timely discharge from hospital, the quality of service received in hospital and the mechanisms in place within the hospital to facilitate timely discharge. It is a delivery contract target, an indicator of older people's services, and a measure of the interface between the NHS and social services. People should receive the right care in the right place at the right time and primary care trusts must ensure, with acute trusts and social services partners, that people move on from the acute environment once they are safe to transfer. The Community Care (Delayed Discharges, etc) Act 2003 facilitates joint working with social services and requires partners to identify the causes of delay, the agency responsible, and the actions required to tackle delays within their local system. Although this is an all-age indicator the vast majority of those delayed are over 75 years. As a result of investment in extra capacity and the introduction of the new Act, delayed transfers of care should be reduced to a minimal level by 2006.

Data source Data Period

LDPR standard collection Financial Year 2003/04
Commission for Health Improvement

18. Diabetic retinopathy screening

Diabetic retinopathy screening

Rationale

The Priorities and Planning Framework states that by 2006, a minimum of 80% of people with diabetes to be offered screening for the early detection (and treatment if needed) of diabetic retinopathy as part of a systematic programme that meets national standards, rising to 100% coverage of those at risk of retinopathy by end 2007.

Data source Data Period

LDPR standard collection To be finalized

DAY 4: (PM): SELECTING M&E DESIGN APPROACH

A key step in the M&E process is the selection of a M&E design approach that fits the context of the health project or program and enables a credible assessment of project implementation and impact.

Performance Monitoring

Performance monitoring is used to provide information on (1) key aspects of how a system or project is operating; (2) whether, and to what extent, pre-specified project objectives are being attained (e.g., numbers of women served by a shelter, increases in cases prosecuted, improved evidence collection); and (3) identification of failures to produce project outputs (this kind of data can be used in managing or redesigning project operations). Performance indicators can also be developed to (4) monitor service quality by collecting data on the satisfaction of those served; and (5) report on project efficiency, effectiveness, and productivity by assessing the relationship between the resources used (project costs and other inputs) and the output and outcome indicators.

If conducted frequently enough and in a timely way, performance monitoring can provide managers with regular feedback that will allow them to identify problems, take timely action, and subsequently assess whether their actions have led to the improvements sought. Performance measures can also stimulate communication about project goals, progress, obstacles, and results among project staff and managers, the public, and other stakeholders. They focus attention on the specific outcomes desired and better ways to achieve them, and can promote credibility by highlighting the accomplishments and value of the project.

Performance monitoring involves identification and collection of specific data on project outputs, outcomes, and accomplishments. Although they may measure subjective factors such as client satisfaction, the data are numeric, consisting of frequency counts, statistical averages, ratios, or percentages. Output measures reflect *internal* activities: the amount of work done within the project or organization. Outcome measures (immediate and longer term) reflect progress towards project goals. Often the same measurements (e.g., number/percent of women who filed for a protection order) may be used for both performance monitoring and impact evaluation. However, unlike impact evaluation, performance monitoring does not make any rigorous effort to determine whether these outcomes were caused by project efforts or by other external events. It centers on answering several key questions:

- Were the intended actions fully implemented?
- Were they implemented on schedule?
- Were they implemented according to design?
- Were the input/output targets reached?
- Were the resources (\$, materials, equipment, etc.) used properly and on schedule?
- Were there any breakdowns in implementation that could affect the outcomes/impacts?
- Were there any drop outs from participation in the program/project? What was there any impact on results?

- Were there “Competing treatments” from other programs?
- Are there any ways in which the implementation of the program/project implementation could be improved?

Design Variations

When projects operate in a number of communities, the sites are likely to vary in mission, structure, the nature and extent of project implementation, primary clients/targets, and timeliness. They may offer somewhat different sets of services, or have identified somewhat different goals. In such situations, it is advisable to construct a "core" set of performance measures to be used by all, and to supplement these with "local" performance indicators that reflect differences. For example, some health projects will collect detailed data on the needs of patients or the history of patient services received, while others will simply have data on the number provided with specific services. Performance indicators need to be constructed so that results can be compared across projects in multi-site projects.

Program Evaluability

Not every program is ready for an impact evaluation (even if it could still profit from a performance monitoring exercise). In addition, every program does not need the same level of evaluation. Since there are never enough evaluation resources to go around, it is important to decide whether impact evaluation is justified and feasible for your project, whether it is likely to provide useful information, and what level of evaluation is needed (monitoring, process evaluation, or impact evaluation). Evaluation planners should ask themselves:

- Does the logic of your diagram stand up to careful scrutiny? Is your proposed link between project activities and objectives reasonable in view of the resources available and findings of earlier research? If project goals are unrealistic or the activities not well grounded in theory and/or prior evidence, then evaluation is not a good investment.
- Is the project operating at its full scope, and is it stable (not just beginning, or not just having changed or being about to change some major aspect of program operations)?
- What kinds of data will be needed, from how many people? What data are likely to be available already from existing data? Evaluations should be designed to maximize the use of available data, as long as these are valid indicators of important concepts and are reliable. Available data may, for example, include government statistics, individual and summary agency records and statistics, and information collected by researchers for other studies. If there are crucial data needs not met with existing data, resources must be available to collect any new data required.
- Are adequate resources and assets available—money, time, expertise, community and government support? Are there any factors that limit or constrain access to these resources?
- Can the evaluation be achieved in a time frame that will permit the findings to be useful in making project and policy decisions?
- To what extent does evaluation information already exist somewhere on the same or a closely related project? The answer to this question can have important implications for action. Any successful previous attempts may yield promising models for replication. Lessons learned from

previous unsuccessful attempts may inform the current effort. If sufficient evidence already exists from previous efforts, the value of a new evaluation may be marginal.

- To what extent are the findings from an evaluation likely to be generalizable to other communities, and therefore useful in assessing whether the project should be expanded to other settings or areas? Are there unique characteristics of the projects to be evaluated that might not apply to most other projects? Project characteristics that are critical to success but are not replicable in other settings reduce the value of any evaluation findings.

Negative or unsatisfactory answers to enough of these questions about a particular project suggest that the project is not ready for impact evaluation, and that major impact evaluation resources would be better spent on projects for which these questions can be answered more positively.

DAY 5 (AM/PM): IMPACT EVALUATION DESIGNS

Impact Evaluation Designs

The following discussion is on the key elements of each design and variations you can consider. The strengths and limitations of each will be summarized as are the general requirements of each in terms of resources such as budget and staff. As you move through these choices, the budget increases as does the extent to which you produce scientifically convincing results. However, as noted below, the best choice is often driven by a consideration of the audience for your results—who wants to know, when do they need to know, what issues do they care about, and what types of information will convince them?

The discussion uses the traditional evaluation terms to describe the people from whom you will be collecting data. Project participants are called "the treatment group" and the services they receive are called "the treatment." Those who do not receive services are called "the control group" (if people are randomly assigned to treatment and control groups) or "the comparison group" (if some method other than random assignment is used to select this group).

Non-Experimental Impact Evaluations

Key Elements. Non-experimental impact evaluations examine changes in outcomes for project participants, or groups that may include project participants (e.g., all women in a particular neighborhood).

Design Variations. You can choose from four primary types of non-experimental design: (1) comparisons of groups before and after treatment; (2) time series designs; (3) panel studies; and (4) cross-sectional comparisons after a treatment has been delivered.

The first two designs are based on analysis of aggregate data—that is, data for *groups*, *not for individuals*. In a **before and after comparison**, outcomes for *groups* of participants that enter the project at a specific time and progress through it over the same time frame are measured before and after an intervention. Your assessment of program impact is inferred from the differences in the average score for the group before and after the services. This simple design is often used to assess whether knowledge, attitudes, or behavior of the group changed after exposure to an intervention. For example, a project focused on training health workers to detect domestic violence might ask whether the average score on knowledge about domestic violence policies increased for your group of participating staff or others after the training compared to the baseline score measured at the start of training. Similarly, you could measure public attitudes or beliefs before and after a public safety campaign aimed at decreasing domestic violence.

A **time series design** is an extension of the before and after design that takes measures of the outcome variables several times before an intervention begins (e.g., once a month for the six months before an intervention starts) and continues to take measures several times after the intervention is in place (e.g., once a month for six months after the intervention). The evaluation tests whether a significant change in direction or level of the outcome occurs at or shortly after the time of the intervention. For example, a project trying to increase community collaboration on health promotion activities could begin collecting information on the number of cross-agency referrals and other collaborative actions every month for the six months before intensive collaboration development efforts begin, and for every month of the two years following the initiation of collaborative work. You could then trace the development of collaborative activity and tie it to events in the community (including the timing of stepped-up efforts to promote collaboration). Time series measures may be collected directly from project participants. However, people also use a time series design based on information from larger groups or units that include but are not restricted to project participants. For example, rates of reported automobile accidents for neighborhoods in which special police

patrols are introduced might be used to assess reductions in automobile accident injuries. A time series design using publicly available data (such as the rate of accidents) should be considered when it is difficult to identify who receives project services, or when the evaluation budget does not support collection of detailed data from project participants. Although statistical techniques can be used to strengthen the usefulness of these designs, it is still difficult to rule out the potential impact of non-project events using this approach.

The next two designs examine data at the individual level (that is, data come from *individuals, not just from groups*). **Cross-sectional comparisons** are based on surveys of *project participants* that you conduct after the project is completed. You can use the data collected with this design to estimate correlations between the outcomes experienced by individuals and differences in the duration, type, and intensity of health services they received. This will let you draw some conclusions about plausible links between outcomes and the types and amount services received within your treatment group. However, you can not draw definitive conclusions about what caused what, because you do not have any type of comparison group that would let you say "it happened for those who got services, but not for those who did not get services." **Panel designs** use repeated measures of the outcome variables for individual participants in a treatment. In this design, outcomes are measured for the same group of project participants, often starting at the time they enter the project and continuing at intervals over time. The design is similar to the "time series" design described earlier, but the data come from individuals, not from groups, and data collection rarely starts *before* the individuals enter the program or receive the intervention.

Considerations/Limitations. Correctly measuring the type and amount of the services received by project participants is critical in non-experimental evaluations. Because the inferences about project impact are based on response to services, differences in the type and amount of service received are critical. The key variations in services need to be spelled out carefully in developing your logic model. Several limitations to non-experimental designs should be noted:

- First, the cross-sectional and panel designs provide only a segment of the "dose-response curve." That is, they only give estimates of the *differences* in impact related to *differences* in the services received. These designs *cannot estimate the full impact of the project compared to no service at all*, unless estimates can be based on other information on the risks of the target population.
- Second, the designs that track participants over time (before and after, panel, and time series) cannot control for the effects of changes that would have occurred anyway without services, or for the effects of other events outside the project's influence.
- Third, the extent to which the results can be assumed to apply to other groups or other settings is limited, because this design provides no information for assessing the extent to which outcomes for those who participated differ from those who might be eligible for services but chose not to participate. For example, if those who came to your training program were those interested in learning more about domestic violence, they might show greater attitude changes after training than a group of trainees required to come by their supervisor. Alternatively, if the project provides intensive counseling only for sexual assault victims who are still showing severe psychological symptoms one or more years after the assault, their gains in mental health might be slower than those of victims whose symptoms (e.g., of fear, anxiety, or depression) abated more quickly.

Practical Issues/Data Collection. Non-experimental designs have several practical advantages. They are relatively easy and inexpensive to conduct. Data from individuals for cross-sectional or panel analyses are often collected routinely by the project at the end (and sometimes beginning) of project participation. When

relying on project records, the evaluator needs to review the available data against the logic model to be sure that adequate information on key variables is already included. If some key data are missing, the evaluator needs to set up procedures for collecting additional data items.

When individual project records are not available, aggregate statistics may be obtained from the project or from other community agencies that have information on the outcomes you care about. The primary problem encountered in using such statistics for assessing impacts is that they may not be available for the specific population or geographic area targeted by the project. Often these routinely collected statistics are based on the general population or geographic areas served by the agency (e.g., the health clinic catchment area). The rates of negative outcomes for the entire set of cases included may well be lower than rates for your target group, if you are trying to serve those with the most severe cases or history of violence. The larger the population or geographical area covered by the statistics, the greater the risk that any effects on program participants will be swamped by the vastly larger number of nonparticipants included in the statistics.

A more expensive form of data collection for non-experimental evaluations is a survey of participants some time after the end of the project. These surveys can provide much needed information on longer term outcomes such as rates of employment or earnings for battered women after leaving the battering situation, or psychological health for sexual assault victims one or more years after the assault. As in any survey research, the quality of the results is determined by response rate rather than by overall sample size, and by careful attention to the validity and reliability of the questionnaire items.

Quasi-Experimental Designs

Key Elements. Quasi-experimental evaluations compare outcomes from project participants to outcomes for comparison groups that do not receive project services. The critical difference between quasi-experimental and experimental designs is that the decision on who participates in the program is not random. Comparison groups are made up of individuals as similar as possible to project participants on factors that could affect the selected outcomes you want to measure. Statistical techniques are then used to control for remaining differences between the groups.

Usually, evaluators use existing groups for comparison—people in the same or similar neighborhoods of the city who did not receive services or those who have similar cases in other neighborhoods. In some situations, selected staff try a new "treatment" (approach to services) while others do not. When selecting a comparison group, you need to be sure that the comparison group is indeed similar to the treatment group on critical factors. If patients are to be served or health personnel are to be trained, those receiving new services should be similar to those who get the existing services.

Design Variations. As just described, the main way to define a comparison group is to find an existing group as similar as possible to the treatment group. The most common variation to the "whole group" approach is called "matching." In matching, the researcher *constructs* a comparison "group" by matching individuals who do not receive treatment to individuals in the treatment group on a selected set of characteristics. This process for constructing a comparison group runs two relatively serious threats to validity. The first is that the groups, while similar at the time of selection, may change over time due to pre-existing characteristics. As a result, changes over time may reflect factors other than the "treatment." The second is that the researcher may have failed to use key variables influencing outcomes the matching process. These variables, which differed between the two groups at the outset, may still cause matched

groups to differ on outcomes for reasons other than the treatment. To do the best possible job on selecting critical variables for matching, you should refer to the background factors which your logic model identifies as likely to influence outcomes. These factors should be used in the match.

Quasi-experimental designs vary in the frequency and timing of collecting data on outcome measures. One makes decisions about the frequency and timing of measurements after assessing the potential threats posed by competing hypotheses that cannot be ruled out by the comparison methodology. In many situations, the strongest designs are those that collect pre-project measures of outcomes and risk factors and use these in the analysis to focus on within-individual changes that occur during the project period. These variables are also used to identify groups of participants who benefit most from the services. One design variation involves additional measurement points (in addition to simple before and after) to measure trends more precisely. Another variation is useful when pre-project data collection (such as administering a test on knowledge or attitudes) might "teach" a sample member about the questions to be asked after the project to measure change, and thus distort the measurement of project impact. This variation involves limiting data collection to the end of the project period for some groups, allowing their post-project answers to be compared with the post-project answers of those who also participated in the pre-project testing.

Considerations/Limitations. Use of non-equivalent control group designs requires careful attention to procedures that rule out competing hypotheses regarding what caused any observed differences on the outcomes of interest.

A major threat in STOP evaluations may be that known as "history" —the risk that unrelated events may affect outcomes. The rapid change in laws, services, and public awareness of violence against women may affect the policies and services available to treatment and comparison groups alike. Changes may occur suddenly in large or small geographic areas, jurisdictions, or service catchment areas. For example, if one clinic begins using a patient advocate successfully, other nearby clinics may adopt the practice or even undertake a more comprehensive project with similar goals.

A second threat to validity is the process of "selection" —the factors that determine who is eligible for, or who chooses to use, services. Some of these factors are readily identified and could be used in selecting the comparison sample, or could be included in the statistical models estimating project impact. For example, if victims who do not speak English are excluded from services either formally or informally, the comparison of outcomes needs to consider English language proficiency as a control variable. Such differences may not be easy to measure during the evaluation.

Practical Issues/Data Collection. It is a challenge to build defenses or "controls" for threats to validity into evaluation designs through the selection of comparison groups and the timing of outcome observations. Even when the comparison group is carefully selected, the researcher cannot be sure that all relevant group differences have been identified and measured accurately. Statistical methods can adjust for such problems and increase the precision with which project effects can be estimated, but they do not fully compensate for the non-random design. Findings need to be interpreted extremely cautiously, and untested alternative hypotheses need to be considered carefully.

Plans for quasi-experimental evaluations need to pay close attention to the problem of collecting comparable information on control group members and developing procedures for tracking them. You may be able to collect data and provide contact information for treatment group members relatively easily because the program and cooperating agencies have continuing contacts with clients, other agencies, and

the community, and have a stake in the outcome of your evaluation. Collecting comparable data and contact information on comparison groups can be difficult. If you collect more complete information for your treatment group than for your comparison group or lose track altogether of more comparison than treatment group members, not only will the evaluation data be incomplete, it will be biased—that is, it will provide distorted and therefore misleading information on project impact. The best way to avoid bias from this problem is to plan tracking procedures and data collection at the start of the evaluation, gathering information from the comparison group members on how they can be located, and developing agreements with other community agencies, preferably in writing, for assistance in data collection and sample member tracking. These agreements are helpful in maintaining continuing contact with your sample in the face of staff turnover at the agencies involved.

Quasi-experimental designs may employ a variety of quantitative and qualitative approaches to gather the data needed to draw conclusions about a project and its impact. Data collection strategies are described below, once we have reviewed all of the options for evaluation design.

Experimental Designs

Key Elements. Experimental designs are considered the "gold standard" in impact evaluation. Experiments require that individuals or groups be assigned at random (by the flip of a coin or equivalent randomizing procedure) to one or more groups prior to the start of project activities. A "treatment" group receives particular services designed to achieve clearly specified outcomes. If several new services are introduced, the experiment can compare multiple treatment groups. A "control" group continues to receive the services in existence prior to the introduction of the new project (either no services or already existing services). The treatment group outcomes are compared to outcomes for alternative treatment groups and/or to a control group to estimate impact. Because chance alone determines who receives the project services, the groups can be assumed to be similar on all characteristics that might affect the outcome measures. Any differences between treatment and control groups, therefore, can be attributed with confidence to the effects of the project.

Design Variations. One design variation is based on a random selection of time periods during which services are provided. For example, new services may be offered on randomly chosen weeks or days. A version of this approach is to use "week on/week off" assignment procedures. Although not truly random, this approach closely approximates random assignment if client characteristics do not vary systematically from week to week. It has the major advantage that project staff often find it easier to implement than making decisions on project entry by the flip of a coin on a case-by-case basis. A second design variation is a staggered start approach in which some members of the target group are randomly selected to receive services with the understanding that the remainder will receive services at a later time (in the case of a school or classroom, the next month, semester, or year). One disadvantage of the staggered start design is that the observations of outcomes are limited to the period between the time the first group completes the project and the second group begins. As a result, it is generally restricted to assessing gains made during participation in relatively short-term projects.

Limitations/Considerations. Although experiments are the preferred design for an impact evaluation on scientific grounds, random assignment evaluations are not always the ideal choice in real life settings. Some interventions are inherently impossible to study through randomized experiments for legal, ethical, or practical reasons. Laws cannot be enforced selectively against a randomly selected subset of offenders or areas in a community. Access to legal protections cannot be curtailed. For example, protection orders

cannot be issued to victims only during selected weeks. Essential services should not be withheld. However, it may be possible to randomly assign alternative services or responses if the relative merits of the alternatives are unknown.

You need to ask yourself whether the results that are likely to be obtained justify the investment. Experiments typically require high levels of resources—money, time, expertise, and support from project staff, government agencies, funders, and the community. Could the answers to evaluation questions—and subsequent decisions on project continuation, expansion, or modification—be based on less costly, less definitive, but still acceptable evaluation strategies? The answer is often "yes."

Practical Issues/Data Collection. Experimental designs run the most risk of being contaminated because of deliberate or accidental mistakes made in the field. To minimize this danger, there must be close collaboration between the evaluation team and the project staff in identifying objectives, setting schedules, dividing responsibilities for record-keeping and data collection, making decisions regarding client contact, and sharing information on progress and problems. Active support of the key project administrators, ongoing staff training, and communication via meetings, conference calls, or e-mail are essential. Failure to adhere to the plan for random assignment is a common problem. Staff are often intensely committed to their clients and will want to base project entry decisions on their perceptions of who needs or will benefit most from the project—although these judgments may not be supported by later research. Thus it is important that the evaluator, not project staff, remain in charge of the allocation to treatment or control group.

As in quasi-experimental evaluations, lack of comparable information for treatment and control group members can be a problem. Experiments generally use both agency records and data collected from individuals through questionnaires and surveys. To assure access to these individuals, quasi-experimental evaluations need to plan for data collection and tracking of sample members at the start of the project and get agreements with agencies and consent procedures with individuals in place early in the process. Along with all other types of impact evaluation, quasi-experimental designs often combine quantitative data with qualitative information gathered through process evaluation in order to understand more about the program when interpreting impacts on participants. Another issue is documenting what parts of the program each participant received. If the project services and content change over time, it may be difficult to determine what level or type of services produced the outcomes. The best strategy is to identify key changes in the project and the timing of changes as part of a process evaluation and use this information to define "types of project" variations in the project experience of different participants for the impact analysis.

Process Analysis

Key Elements

Process evaluations rarely vary in basic design. Most involve a thorough documentation and analysis of activities of the program. A good process analysis design is guided by a set of core questions: Is the project model being implemented as specified and, if not, how do operations differ from those initially planned? Does the program have unintended consequences and unanticipated outcomes and, if so, what are they and who is affected? What is the view of the project from the perspectives of staff, participants, and the community? The answers to these questions are useful in providing guidance to policy makers and project planners interested in identifying key project elements and in generating hypotheses about project impact that can be tested in impact analyses.

Design Variations

Process evaluations vary in the number of projects or sites included. Most process evaluations focus on a single project or site. However, some undertake comparative process analysis. Comparative process analysis requires that observations, interviews, and other data collection strategies be structured in advance around a set of questions or hypotheses about elements of implementation believed to be critical to project success. Comparative process analysis allows the evaluation to make assessments about alternative strategies and is useful in generalizing the findings to other settings or jurisdictions. This strategy is used to assess which approach is most successful in attaining goals shared by all when competing models have emerged in different locations. It requires purposely selecting sites to represent variations in elements or types of projects, careful analysis of potential causal models, and the collection of qualitative data to elaborate the causal links at each site.

Most design uncertainties in process evaluation involve deciding what information will be collected, from whom and how. Process evaluation can be based solely on qualitative data. However, qualitative data are usually combined with quantitative data on services produced, resources used, and outcomes achieved. Qualitative data collection strategies used in process evaluation include semi-structured interviews with those involved in project planning and operations; focus groups with project planners, staff, or participants; and researcher observations of project activities

Practical Issues

In a process evaluation, it is often difficult to decide on what information is truly key to describing program operations and what information is simply extraneous detail. In selecting relevant data and posing questions about program operations, the evaluator needs to refer carefully to the logic model prepared at the start of the project, although it is permissible and important in process evaluation to revise the original logic model in light of findings during the evaluation.

Analysis of qualitative data requires considerable substantive knowledge on the part of the evaluator. The evaluator needs to be familiar with similar projects, respondents, and responses, and the context in which the project is operating. Your evaluator will need to be able to understand the project's historical and political context as well as the organizational setting and culture in which services are delivered. At the same time, the evaluator needs to maintain some objectivity and separation from project management in order to be able to make an unbiased assessment of whether responses support or refute hypotheses about the way the project works and the effects it has.

Collecting qualitative data also requires skilled researchers who are experienced in interviewing and observing. Data must be carefully recorded or taped. Notes on contextual factors and interim hypotheses need to be recorded as soon as possible after data collection. When using interview guides or semi-structured interview protocols, interviewers must be trained to understand the intent of each question, the possible variety of answers that respondents might give, and ways to probe to ensure that full information about the issues under investigation is obtained.

Exercise: Participants discuss what they see as the factors affecting their choice of a M&E design. Trainers serve as resource persons

Identifying Alternative Explanations for M&E Results

Identifying alternative explanations (i.e., counterfactuals) is at the core of evaluation design. It is, however, quite tricky to net out the program impact from the counterfactual conditions that are likely to be influenced

by contemporaneous events, selection bias, and contamination. Details of a program can and should affect the choice of the evaluation method and how it is carried out. In particular, the evaluation should be designed to minimize any risk that the study itself might compromise the program by altering the program's delivery in some fundamental way, changing the type of individual who would be served, or changing the behavior of the members of the control group.

The following operational issues of an evaluation study might compromise the objective of the program and thus might make it difficult to conduct the study or to draw any inference once it is conducted.

- **Altering the services delivered:** Sometimes the evaluation study can provide new information that might alter the program's delivery in some fundamental way thus affecting outcomes. Also, the study can overburden program staff with research requirements, reducing their time and energy to deliver services.
- **Changing Target Population:** Research requirements can skew selection of participants so that the population served by the pilot program is different from the intended recipients of the program itself. For example, individuals who are willing to participate in the evaluation study may differ from individuals who are simply willing to receive services.
- **Contamination of the control group** due to (a) attrition, (b) migration between treated and untreated groups and (c) information leakage between the two groups and (d) unintended receipt of program services by the group.
- **Constraints on time and resources:** Random assignment and prospective comparison group designs (including prospective reflexive studies) need a baseline survey. So, they are more expensive than retrospective comparison group designs which do not have to allow time for building up a sample for treatment and for assessing post program effects. This is an important issue to take into account when designing an evaluation. Also, given the constraints, there is always a tension between conflicting objectives- investigating short run effects versus long-term effects of the intervention.

Threats to Validity

Besides the operational problems cited above, there are some general problems with drawing conclusions from empirical evaluation studies. These problems pose threats to internal validity as well as external validity. Internal validity refers to whether one can validly draw the inference that within the context of the study the differences in the dependent variables were caused by differences in the relevant explanatory variables. External validity deals with whether effects found in an experiment can be generalized to different individuals, contexts, and outcomes.

The threats to internal validity could be due to several reasons:

- a. Omitted variables, for example, events, other than the experimental treatment, occurring between pre-intervention and post intervention observations that provide alternative explanations for the results.
- b. Trends in outcomes, when processes within the units of observations produce changes as a function of the passage of time per se, such as inflation, aging, and wage growth.

- c. Misspecified variances, causing overstatement of the significance of statistical tests due to effects such as the omission of group error terms that indicate that outcomes for individual units are correlated.
- d. Mismeasurement, that is, changes in definitions or survey methods that may produce changes in the measured variables.
- e. Political economy, for example, endogeneity of policy changes due to governmental responses to variables associated with past or expected future outcomes.
- f. Simultaneity, that is, endogeneity of explanatory variables due to their joint determination with outcomes.
- g. Selection bias, that is, assignment of observations to treatment groups in a manner that leads to correlation between assignment and outcomes in the absence of treatment.
- h. Attrition, which is the differential loss of respondents from treatment and comparison groups.
- i. Dropouts of some of the experimental treatment group members from the social program under study prior to receiving some or all of the treatment. This is different from the attrition problem as persons drop out of the program, but not out of the experimental data.
- j. Omitted interactions, for example, exclusion of interactions such as differential trends in treatment and control groups or omitted variables that change in different ways for treatment and control groups.

The threats to external validity are just the possibility that there are important interactions between the treatment and individual characteristics, location, or time such that results from an evaluation study can not be generalized to different individuals, contexts, and outcomes. These interactions are as follows:

- Interaction of selection and treatment: Unrepresentative responsiveness of the treated population. The treatment group may not be representative of certain population, or the treatment may be different from that which one would like to examine.
- Interaction of setting and treatment: The effect of the treatment may differ across geographic or institutional settings.
- Finally, interaction of history and treatment: The effect of the treatment may differ across time periods.

Exercise: Participants discuss what threats are most likely to be most important in their particular areas of health services provision. How do they plan to take into account for these threats in the M&E planning?

DAY 6 (AM/PM): COLLECTING DATA ON M&E INDICATORS

Data collection for M&E vary along a continuum. At the one end of this continuum are methods relying on random sampling; structured data collection instruments that fit diverse experiences into predetermined response categories; and statistical data analysis. These methods, generally associated with quantitative research, produce results that are easy to summarize, compare, and generalize.

At the other end of the continuum are methods typically associated with qualitative research. These qualitative methods are characterized by the following attributes:

- they tend to be open-ended and have less structured protocols (i.e., researchers may change the data collection strategy by adding, refining, or dropping techniques or informants)
- they rely more heavily on iterative interviews; respondents may be interviewed several times to follow up on a particular issue, clarify concepts or check the reliability of data
- they use triangulation to increase the credibility of their findings (i.e., researchers rely on multiple data collection methods to check the authenticity of their results)
- generally their findings are not generalizable to any specific population, rather each case study produces a single piece of evidence that can be used to seek general patterns among different studies of the same issue.

In between the two extremes, there is a number of possible evaluation methodologies combining different aspects (sample design, research protocol, data collection and data analysis) of the quantitative and qualitative approaches.

Evaluations can also rely on participatory methods. These tend to be closer to the qualitative than to the quantitative research approach. However, not all qualitative methods are participatory, and inversely, many participatory techniques can be quantified. The participatory approach is very much action-oriented. Thus, stakeholders themselves are responsible for collecting and analyzing the information, and for generating recommendations for change. The role of an outside evaluator is to facilitate and support this learning process.

By combining these different approaches, one can enrich the design as well as the interpretation or explanation of outcomes measured by the evaluation.

Qualitative Methods

Qualitative methods for data collection play an important role in impact evaluation by providing information useful to understand the processes behind observed results and assess changes in people's perceptions of their well-being. Furthermore qualitative methods can be used to improve the quality of survey-based quantitative evaluations by helping generate evaluation hypothesis; strengthening the design of survey questionnaires and expanding or clarifying quantitative evaluation findings.

The qualitative methods most commonly used in evaluation can be classified in three broad categories: In-depth interviewing entails asking questions, listening to and recording the answers, and then posing additional questions to clarify or expand on a particular issue. Questions are open-ended and respondents

are encouraged to express their own perceptions in their own words. In-depth interviewing aims at understanding the beneficiaries' view of a program, their terminology and judgments.

There are three basic approaches to in-depth interviewing that differ mainly in the extent to which the interview questions are determined and standardized beforehand: the informal conversational interview; semi-structured interview; and the standardized open-ended interview. Each approach serves a different purpose and has different preparation and instrumentation requirements.

The **informal conversational interview** relies primarily on the spontaneous generation of questions in the natural flow of an interaction. This type of interview is appropriate when the evaluator wants to maintain maximum flexibility to be able to pursue questioning in whatever direction appears to be appropriate, depending on the information that emerges from observing a particular setting, or from talking to one or more individuals in that setting. Under these circumstances, it is not possible to have a predetermined set of questions. The strength of this approach is that the interviewer is flexible and highly responsive to individual differences, situational changes and emerging new information. The weakness is that it may generate less systematic data that is difficult and time consuming to classify and analyze.

Semi-structured interviews involve the preparation of an interview guide that lists a pre-determined set of questions or issues that are to be explored during an interview. This guide serves as a checklist during the interview and ensures that basically the same information is obtained from a number of people. Yet, there is a great deal of flexibility. The order and the actual working of the questions is not determined in advance. Moreover, within the list of topic or subject areas, the interviewer is free to pursue certain questions in greater depth. The advantage of the interview guide approach is that it makes interviewing of a number of different persons more systematic and comprehensive by delimiting the issues to be taken up in the interview. Logical gaps in the data collected can be anticipated and closed, while the interviews remain fairly conversational and situational. The weakness of this approach is that it does not permit the interviewer to pursue topics or issues of interest that were not anticipated when the interview guide was elaborated. Also, interviewer flexibility in wording and sequencing questions may result in substantially different responses from different persons, thus reducing comparability.

The **standardized open-ended interview** consists of a set of open-ended questions carefully worded and arranged in advance. The interviewer asks the same questions to each respondent with essentially the same words and in the same sequence. This type of interview may be particularly appropriate when there are several interviewers and the evaluator wants to minimize the variation in the questions they pose. It is also useful when it is desirable to have the same information from each interviewee at several points in time or when there are time constraints for data collection and analysis. Standardized open-ended interviews allow the evaluator to collect detailed data systematically and facilitate comparability among all respondents. The weakness of this approach is that it does not permit the interviewer to pursue topics or issues that were not anticipated when the interview instrument was elaborated. Also, standardized open-ended interviews limits the use of alternative lines of questioning with different people depending on their particular experiences. This reduces the extent to which individual differences and circumstances can be fully incorporated in the evaluation.

Interviews with Individual Respondents

A common type of individual respondent interview is the **key informant interview**. A key informant is an individual, who as a result of their knowledge, previous experience or social status in a community has access to information valuable for the evaluator such as insights about the functioning of society, their problems and needs. Key informants are a source of information which can assist in understanding the context of a program or project, or clarifying particular issues or problems. However, since the selection of key informants is not random, the issue of bias always arises. Another difficulty of this method lies in separating the informants' potential partiality to form a balanced view of the situation.

Group Interviews

Interviews with a group of individuals can take many different forms depending on the purpose they serve, the structure of the questions, the role of the interviewer and the circumstances under which the group is convened. Some of the group interview types relevant to evaluation are: focus groups, community interviews and spontaneous group interviews.

Focus group interviews are interviews with small groups of relatively homogeneous people with similar background and experience. Participants are asked to reflect on the questions asked by the interviewers, provide their own comments, listen to what the rest of the group have to say and react to their observations. The main purpose is to elicit ideas, insights and experiences in a social context where people stimulate each other and consider their own views along with the views of others. Typically, these interviews are conducted several times with different groups so that the evaluator can identify trends in the perceptions and opinions expressed. The interviewer acts as facilitator introducing the subject, guiding the discussion, cross-checking each other comments and encouraging all members to express their opinions. One of the main advantages of this technique is that participant interaction helps weed out false or extreme views, thus providing a quality control mechanism. This, however, requires a skillful facilitator to ensure an even participation from all members.

Community interviews are conducted as public meetings in which the whole community is consulted. Typically, these interviews involve a set of factually-based fairly close-ended questions. Once the interviewers pose the question, the group will interact to get a consensus around an answer. Interviewing the community as a whole can provide valuable information on how well a project is working. The major weakness of these methods is that participation may be limited to a few high status members of the community or that community leaders may use the forum to seek consensus on their own views and preferences.

Observational Methods

Firsthand observation of a program is another important source of qualitative data for evaluation. The main purpose of observational evaluation is to obtain a thorough description of the program including program activities, participants and the meaning they attach to the program. It involves careful identification and accurate description of relevant human interactions and processes.

There are several advantages to observational fieldwork for evaluation purposes:

- it provides a better understanding of the context in which program activities occur;
- it allows to be aware of important things program participants may ignore or omit willingly or unwillingly in an interview;

- it permits the evaluator to present a more comprehensive view of the program by combining his own as well as others perceptions;
- it helps understand and interpret the program by providing personal knowledge and direct experience.

Observational evaluation is a hard work that requires a skilled, trained and competent evaluator to ensure good quality data. There are a number of variations in observational methods. The most fundamental difference among them refers to the role of the evaluation observer either as a full program participant, a detached spectator or somewhere in between.

Participant observation is at one end of the participation spectrum and consists of the evaluation observer becoming a member of the community or population being studied. The researcher participates in activities of the community, observes how people behave and interact with each other and outside organizations. The evaluator tries to become accepted as a neighbor or participant rather than as an outsider. The purpose of such participation is not only to see what is happening but to feel what it is like to be part of the group. The extent to which this is possible depends on the characteristics of program participants, the type of questions being studied and the socio-political context of the setting. The strength of this approach is that the researcher is able to experience and presumably better understand any project impacts. The main weakness is that it is likely to alter the behavior that is being observed. In addition, ethical issues may arise if the participant observer misrepresents himself/herself in order to be accepted by the community being studied.

Direct observation tends to be at the other end of the participation spectrum. It involves the systematic noting and recording of activities, behaviors and physical objects in the evaluation setting as an unobtrusive observer. It can often be a rapid and economical way of obtaining basic socio-economic information on households or communities. The main advantage of this method is that if participants are not aware that they are being observed, then they are less likely to change their behavior and compromise the validity of the evaluation.

It is important to remember that there is a great deal of variation between the two extremes and that the extent of participation can change over time. For example, the evaluator may begin the observation as an outsider and gradually become a participant as the study evolves.

Document Review

Evaluators may supplement observational fieldwork and interviewing with gathering and analyzing documentary material generated by a program such as laws, regulations, contracts, correspondence, memoranda and routine records on services and clients. These kinds of documents are a useful source of information on program activities and processes, and they can generate ideas for questions that can be pursued through observation and interviewing. In addition, program documents can provide valuable information that may not be accessible by other means. For example, they can provide information about things that the evaluator cannot observe because they took place before the evaluation began, they were part of private interchanges in which the evaluator did not participate or they reflect plans that have not been realized in actual program performance.

A major advantage of this method is the documents were generated contemporaneously with the events they refer to. Hence, they are less likely to be subject to memory decay or memory distortion compared with data obtained from an interview. However, an important disadvantage is that they may be subject to selective-deposit or selective-survival bias.

Participatory Methods

The participatory approach to evaluation is aimed at promoting action and community-level change. It tends to overlap more with qualitative than with quantitative methods. However, not all qualitative methods are participatory, and inversely, many participatory techniques can be quantified.

As with qualitative methods, participatory evaluation ensures that the perspectives and insights of all stakeholders, beneficiaries as well as project implementers, are taken into consideration. However, the participatory approach is very much action-oriented. The stakeholders themselves are responsible for collecting and analyzing the information, and for generating recommendations for change. The role of an outside evaluator is to facilitate and support this learning process. Participatory M&E develops ownership by placing a strong emphasis on building the capacity and commitment of all stakeholders to reflect, analyze, and take responsibility for implementing any changes they recommend.

Typically, participatory methods have been used to learn about local-level conditions and local people's perspectives and their priorities during project appraisal. But one can go further, and use participatory methods not only at project formulation stage, but throughout the duration of the project, and especially for evaluating how the poor perceived the benefits from the project. Participatory monitoring and evaluation (PM&E) is an important management tool that provides task managers with quick feedback on project effectiveness during implementation. This has become increasingly important as development interventions move away from "blueprint projects" toward the more flexible planning which enables projects to learn and adapt on-the-ground.

There are many different participatory information collection and analysis tools. Most of these are not inherently M&E tools, but can be used for a range of purposes ranging from project planning and community mobilization through M&E depending on the way they are employed. As with all participatory approaches, the key to success is to be flexible and innovative in the use of appropriate tools and methods, and to be willing to adapt to local circumstances.

This site provides descriptions of three participatory methodologies and their associated tools and techniques which are commonly used in participatory M&E:

Beneficiary Assessment

Beneficiary Assessment (BA) is a consultative methodology used in evaluations (and other stages of the project cycle) to gain insights into the perceptions of beneficiaries regarding a project or policy. The overall objective of a BA is to make the voices of beneficiaries and other local-level stakeholders heard by those managing a project or formulating policy.

The focus of BAs is on obtaining systematic qualitative information, including subjective opinions, to complement the data from quantitative evaluations. Wherever possible, BA results are quantified and tabulated. Moreover, sample sizes are selected with credibility in mind. Although BA results are not usually conducive to statistical analysis, they are based on more than just anecdotal information. The systematic nature of BAs also enhances the reliability of the findings due to the combination of techniques used to

gather information. Such techniques allow for crosschecking of responses, and a reasonable assessment of the extent to which opinions expressed by respondents represent widely held views in their community. However, the actual techniques used and the BA process itself will depend on the topic and circumstances of the work.

In addition to generating descriptive information, BAs are designed to produce recommendations, as suggested by those consulted, for changes to the current or planned policies and programs. This action-oriented nature of BA work requires that the results be produced with a minimum of delay after completion of fieldwork so that the necessary adjustments to projects or policies can then be identified and undertaken.

The most common application of BA techniques has been in projects with a service delivery component where it is especially important to gauge user demand and satisfaction. During implementation, BAs can provide feedback for monitoring purposes and for reorientation of the project. Towards the end of the project, BAs can also complement technical and financial evaluations, as well as survey-based impact evaluations with the views of the beneficiaries themselves.

The primary audience of BA findings are decision makers and managers of the development activity. For this reason, special efforts are made to seek the involvement of these decision makers in the BA process from the design stage to the review and final presentation of the results.

Usually, BAs make use of three qualitative methods of information gathering:

- semi-structured individual interviews
- focus group discussions
- participant observation

Semi-structured interviews provide the bulk of the findings. They are meant to be quantified -and hence the sample must be large enough and representative. Focus group interviews and participant observation are done primarily for illustration and contextual background. and need not conform the same standards of representativeness.

The quality and effectiveness of BAs depends heavily on the training and preparedness of the field workers and the appropriate supervision and monitoring of their work. Where field workers are unclear about the kind of information required for the evaluation, the common tendency is to collect lengthy, descriptive and very detailed information on individual cases, rather than focusing only on the relevant topics. For this reason, there should be at least one opportunity to review the preliminary findings and methods, preferably midway through the fieldwork, so these kinds of problems can be addressed in time to reorient the field workers' work.

Another limitation seen in some BAs is the failure to ensure active participation by key decision makers throughout the process. In this case, even if the findings are of good quality and highly relevant, they are unlikely to generate much impact. Without a sense of ownership, decision makers may not accept the findings, particularly if they are somewhat controversial and critical of the project or policy concerned. This caveat applies to all evaluation work regardless of the type of approach or technique used.

Participatory Rural Appraisal (PRA)

Participatory Rural Appraisal (PRA) comprises a set of techniques aimed at shared learning between local people and outsiders. The term itself is misleading because more and more PRA is being used not only in rural settings, and not only for project appraisal, but throughout the project cycle, as well as for research studies. Indeed, the term PRA is one of many labels for similar participatory assessment approaches, the methodologies of which overlap considerably. It is probably more useful to consider the key principles behind PRA and its associated techniques, rather than the name per se, when assessing its appropriateness to a particular situation.

There are five key principles that form the basis of any PRA activity no matter what the objectives or setting:

1. **Participation.** PRA relies heavily on participation by the communities, as the method is designed to enable local people to be involved, not only as sources of information, but as partners with the PRA team in gathering and analyzing the information.
2. **Flexibility.** The combination of techniques that is appropriate in a particular development context will be determined by such variables as the size and skill mix of the PRA team, the time and resources available, and the topic and location of the work.
3. **Teamwork.** Generally, a PRA is best conducted by a local team (speaking the local languages) with a few outsiders present, a significant representation of women, and a mix of sector specialists and social scientists, according to the topic.
4. **Optimal Ignorance.** To be efficient in terms of both time and money, PRA work intends to gather just enough information to make the necessary recommendations and decisions.
5. **Systematic.** As PRA-generated data is seldom conducive to statistical analysis (given its largely qualitative nature and relatively small sample size), alternative ways have been developed to ensure the validity and reliability of the findings. These include sampling based on approximate stratification of the community by geographic location or relative wealth, and cross-checking, that is using a number of techniques to investigate views on a single topic (including through a final community meeting to discuss the findings and correct inconsistencies).

PRA offers a "basket of techniques" from which those most appropriate for the project context can be selected. The central part of any PRA is semi-structured interviewing. While sensitive topics are often better addressed in interviews with individuals, other topics of more general concern are amenable to focus group discussions and community meetings.

PRA involve some risks and limitations. Many of them are not unique to this method but are inherent in any research method that aims to investigate local conditions. One of the main problems is the risk of raising expectations. This may be impossible to avoid, but can be minimized with careful and repeated clarification of the purpose of the PRA and the role of the team in relation to the project, or government, at the start of every interview and meeting. Trying to use PRA as a standard survey to gather primarily quantitative data, using large sample sizes, and a questionnaire approach could greatly compromise the quality of the work

and the insights produced. And, if the PRA team is not adequately trained in the methodology before the work begins, there is often a tendency to use too many different techniques, some of which are not relevant to the topic at hand. In general, when a training element is involved, there will be a trade-off between the long-term objective of building the capacity of the PRA team and getting good quality results in their first experience of using the methodology.

Furthermore, one common problem is that insufficient time is allowed for the team to relax with the local people, to listen to them, and to learn about the more sensitive issues under consideration. Rushing will also often mean missing the views of the poorest and least articulate members of the communities visited. The translation of PRA results into a standard evaluation report poses considerable challenges, and individuals unfamiliar with participatory research methods may raise questions about the credibility of the PRA findings.

These methods can be used alone or combined in a single evaluation. They represent only a small sample of the vast range of participatory techniques that can be used for M&E.

It should be noted that none of these participatory methods is intended to be a replacement for good quality survey work. Indeed, they are often used in conjunction with other methods. For example, the findings from a preliminary study using PRA or SARAR techniques can usefully give direction and focus to a subsequent survey-based evaluations. In turn, the survey can verify and quantify the qualitative findings from participatory evaluations and be applied on a larger scale. Participatory evaluations done after quantitative surveys can verify or challenge survey findings, and can go some way toward explaining the information collected by the quantitative survey-based evaluations.

Integrated approach

There is a growing acceptance of the need for integrating the different approaches to evaluation. Impact evaluations using survey data from statistically representative samples may be better suited to assessing causality by using econometric methods or reaching generalizable conclusions. However, qualitative and participatory methods allow the in-depth study of selected issues, cases, or events and can provide critical insights into beneficiaries' perspectives, the dynamics of a particular reform, or the reasons behind certain results observed in a quantitative analysis.

Integrating quantitative, qualitative and participatory methods can often be the best vehicle for meeting the program's information needs. For example, qualitative methods can be used to inform the evaluation questions and the questionnaire design, as well as to analyze the social, economic, and political context within which a program or policy takes place. Similarly, quantitative methods can be used to inform qualitative data collection strategies, including sample design, and, apply statistical analysis to control for household characteristics and the socio-economic conditions of different study areas, thereby eliminating alternative explanations of the observed outcomes.

There are a number of benefits of using integrated approaches in impact evaluations including the following:

- Consistency checks can be built in through the use of triangulation procedures that permit two or more independent estimates to be made for key variables (such as income, opinions about projects, reasons for using or not using public services, and specific impact of a project).

- Different perspectives can be obtained. For example, although researchers may consider income or consumption to be the key indicators of household welfare, case studies may reveal that women are more concerned about vulnerability (defined as the lack of access to social support systems in times of crises), powerlessness, or exposure to violence.
- Analysis can be conducted on different levels. Survey methods can provide good estimates of individual, household, and community-level welfare, but they are much less effective for analyzing social processes (social conflict, reasons for using or not using services, and so on) or for institutional analysis (how effectively health, education, credit, and other services operate and how they are perceived by the community). There are many qualitative methods designed to analyze issues such as social process, institutional behavior, social structure, and conflict.
- Opportunities can be provided for feedback to help interpret findings. Survey reports frequently include references to apparent inconsistencies in findings or to interesting differences between communities or groups that cannot be explained by the data. In most quantitative research, once the data collection phase is completed it is not possible to return to the field to check on such questions. In many cases the data analyst has to make an arbitrary decision as to whether a household or community that reports conditions that are significantly above or below the norm should be excluded (on the assumption that it reflects a reporting error) or the figures adjusted. The greater flexibility of qualitative research means that it is often possible to return to the field to gather additional data. Thus allowing a rapid follow-up in the field to check on these cases.

Exercise: Participants discuss what types of data collection methods seem to fit their health service program context. What type of data do they think they will be relying upon in their M&E exercise? Will they be able to collect quantitative

DAY 7 (AM): BASIC SAMPLING ISSUES

What is a Sample?

A sample is a finite part of a statistical population whose properties are studied to gain information about the whole (Webster, 1985). When dealing with people, it can be defined as a set of respondents (people) selected from a larger population for the purpose of a survey.

A population is a group of individuals, persons, objects, or items from which samples are taken for measurement for example a population of presidents or professors, books or students.

What is sampling? Sampling is the act, process, or technique of selecting a suitable sample, or a representative part of a population for the purpose of determining parameters or characteristics of the whole population.

What is the Purpose of Sampling?

To draw conclusions about populations from samples, we must use inferential statistics which enables us to determine a population's characteristics by directly observing only a portion (or sample) of the population. We obtain a sample rather than a complete enumeration (a census) of the population for many reasons. Obviously, it is cheaper to observe a part rather than the whole, but we should prepare ourselves to cope with the dangers of using samples. In this tutorial, we will investigate various kinds of sampling procedures. Some are better than others but all may yield samples that are inaccurate and unreliable. We will learn how to minimize these dangers, but some potential error is the price we must pay for the convenience and savings the samples provide.

There would be no need for statistical theory if a census rather than a sample was always used to obtain information about populations. But a census may not be practical and is almost never economical. There are six main reasons for sampling instead of doing a census. These are; -Economy -Timeliness -The large size of many populations -Inaccessibility of some of the population -Destructiveness of the observation - accuracy.

The Economic Advantage of Using a Sample in Research

Obviously, taking a sample requires fewer resources than a census. For example, let's use one of our most prestigious universities, Cornell University in New York State. One might want to know what all the students at Cornell think about the quality of teaching they receive. You know that all the students are different so they are likely to have different perceptions and you believe you must get all these perceptions so you decide because you want an in-depth view of every student, you will conduct personal interviews with each one of them and you want the results in 20 days only, let us assume this particular time you are doing your research. Cornell University has only 20,000 students and those who are helping are so fast at the interviewing art that together you can interview at least 10 students per person per day in addition to your 18 credit hours of course work. You will require 100 research assistants for 20 days and since you are paying them minimum wage of \$5.00 per hour for ten hours (\$50.00) per person per day, you will require \$100000.00 just to complete the interviews, analysis will just be impossible. You may decide to hire additional assistants to help with the analysis at another \$100000.00 and so on assuming you have that amount on your account.

As unrealistic as this example is, it does illustrate the very high cost of census. For the type of information desired, a small wisely selected sample of Cornell students can serve the purpose. You don't even have to hire a single assistant. You can complete the interviews and analysis on your own. Rarely does a circumstance require a census of the population, and even more rarely does one justify the expense.

The Time Factor

A sample may provide you with needed information quickly. For example, you are a Doctor and a disease has broken out in a village within your area of jurisdiction, the disease is contagious and it is killing within hours nobody knows what it is. You are required to conduct quick tests to help save the situation. If you try a census of those affected, they may be long dead when you arrive with your results. In such a case just a few of those already infected could be used to provide the required information.

The Very Large Populations

Many populations about which inferences must be made are quite large. For example, consider the population of high school seniors in United States of America, a group numbering 4,000,000. The responsible agency in the government has to plan for how they will be absorbed into the different departments and even the private sector. The employers would like to have specific knowledge about the student's plans in order to make compatible plans to absorb them during the coming year. But the big size of the population makes it physically impossible to conduct a census. In such a case, selecting a representative sample may be the only way to get the information required from high school seniors.

The Partly Accessible Populations

There are some populations that are so difficult to get access to that only a sample can be used. Like people in hospitals, like crashed airplanes in the deep seas, presidents, etc. The inaccessibility may be economic or time related. Like a particular study population may be so costly to reach like the population of planets that only a sample can be used. In other cases, a population of some events may be taking too long to occur that only sample information can be relied on. For example natural disasters like a flood that occurs every 100 years or take the example of the flood that occurred in Noah's days. It has never occurred again.

Accuracy and Sampling

A sample may be more accurate than a census. A sloppily conducted census can provide less reliable information than a carefully obtained sample.

Bias and Error in Sampling

A sample is expected to mirror the population from which it comes, however, there is no guarantee that any sample will be precisely representative of the population from which it comes. Chance may dictate that a disproportionate number of untypical observations will be made like for the case of testing fuses, the sample of fuses may consist of more or less faulty fuses than the real population proportion of faulty cases. In practice, it is rarely known when a sample is unrepresentative and should be discarded.

Sampling Error

What can make a sample unrepresentative of its population? One of the most frequent causes is sampling error.

Sampling error comprises the differences between the sample and the population that are due solely to the particular units that happen to have been selected.

For example, suppose that a sample of 100 American women are measured and are all found to be taller than six feet. It is very clear even without any statistical prove that this would be a highly unrepresentative sample leading to invalid conclusions. This is a very unlikely occurrence because naturally such rare cases are widely distributed among the population. But it can occur. Luckily, this is a very obvious error and can be detected very easily.

The more dangerous error is the less obvious sampling error against which nature offers very little protection. An example would be like a sample in which the average height is overstated by only one inch or two rather than one foot which is more obvious. It is the unobvious error that is of much concern. There are two basic causes for sampling error. One is chance: That is the error that occurs just because of bad luck. This may result in untypical choices. Unusual units in a population do exist and there is always a possibility that an abnormally large number of them will be chosen. The main protection against this kind of error is to use a large enough sample.

The second cause of sampling is sampling bias. Sampling bias is a tendency to favor the selection of units that have particular characteristics. Sampling bias is usually the result of a poor sampling plan. The most notable is the bias of non response when for some reason some units have no chance of appearing in the sample. A means of selecting the units of analysis must be designed to avoid the more obvious forms of bias. An example would be where you would like to know the average income of some community and you decide to use the telephone numbers to select a sample of the total population in a locality where only the rich and middle class households have telephone lines. You will end up with high average income which will lead to the wrong policy decisions.

Non Sampling Error (Measurement Error)

The other main cause of unrepresentative samples is non sampling error. This type of error can occur whether a census or a sample is being used. Like sampling error, non sampling error may either be produced by participants in the statistical study or be an innocent by product of the sampling plans and procedures.

A non sampling error is an error that results solely from the manner in which the observations are made. The simplest example of non sampling error is inaccurate measurements due to malfunctioning instruments or poor procedures. For example, consider the observation of human weights. If persons are asked to state their own weights themselves, no two answers will be of equal reliability. The people will have weighed themselves on different scales in various states of poor calibration. An individual's weight fluctuates diurnally by several pounds, so that the time of weighing will affect the answer. The scale reading will also vary with the person's state of undress. Responses therefore will not be of comparable validity unless all persons are weighed under the same circumstances.

Biased observations due to inaccurate measurement can be innocent but very devastating. A story is told of a French astronomer who once proposed a new theory based on spectroscopic measurements of light emitted by a particular star. When his colleagues discovered that the measuring instrument had been contaminated by cigarette smoke, they rejected his findings.

In surveys of personal characteristics, unintended errors may result from: -The manner in which the response is elicited -The social desirability of the persons surveyed -The purpose of the study -The personal biases of the interviewer or survey writer

The Interviewer Effect

No two interviewers are alike and the same person may provide different answers to different interviewers. The manner in which a question is formulated can also result in inaccurate responses. Individuals tend to provide false answers to particular questions. For example, some people want to feel younger or older for some reason known to themselves. If you ask such a person their age in years, it is easier for the individual just to lie to you by over stating their age by one or more years than it is if you asked which year they were born since it will require a bit of quick arithmetic to give a false date and a date of birth will definitely be more accurate.

The Respondent Effect

Respondents might also give incorrect answers to impress the interviewer. This type of error is the most difficult to prevent because it results from outright deceit on the part of the respondent. It is important to acknowledge that certain psychological factors induce incorrect responses and great care must be taken to design a study that minimizes their effect.

Knowing the Study Purpose

Knowing why a study is being conducted may create incorrect responses. A classic example is the question: What is your income? If a government agency is asking, a different figure may be provided than the respondent would give on an application for a home mortgage. One way to guard against such bias is to camouflage the study's goals; another remedy is to make the questions very specific, allowing no room for personal interpretation. For example, "Where are you employed?" could be followed by "What is your salary?" and "Do you have any extra jobs?" A sequence of such questions may produce more accurate information.

Induced Bias

Finally, it should be noted that the personal prejudices of either the designer of the study or the data collector may tend to induce bias. In designing a questionnaire, questions may be slanted in such a way that a particular response will be obtained even though it is inaccurate. For example, an agronomist may apply fertilizer to certain key plots, knowing that they will provide more favorable yields than others. To

protect against induced bias, advice of an individual trained in statistics should be sought in the design and someone else aware of search pitfalls should serve in an auditing capacity.

Selecting the Sample

The desirability of a sampling procedure depends on both its vulnerability to error and its cost. However, economy and reliability are competing ends, because, to reduce error often requires an increased expenditure of resources. Of the two types of statistical errors, only sampling error can be controlled by exercising care in determining the method for choosing the sample. The previous section has shown that sampling error may be due to either bias or chance. The chance component (sometimes called random error) exists no matter how carefully the selection procedures are implemented, and the only way to minimize chance sampling errors is to select a sufficiently large sample (sample size is discussed towards the end of this tutorial). Sampling bias on the other hand may be minimized by the wise choice of a sampling procedure.

Types of Samples

There are three primary kinds of samples: the convenience, the judgment sample, and the random sample. They differ in the manner in which the elementary units are chosen.

1. The Convenience Sample

A convenience sample results when the more convenient elementary units are chosen from a population for observation. In other words, the convenience sample is comprised of people it was very easy to contact, they were readily available, regardless of their characteristics.

2. The Judgment Sample

A judgment sample is obtained according to the discretion of someone who is familiar with the relevant characteristics of the population. People are picked because they exhibit certain desirable traits, such as living in a certain area or being from a certain ethnic group.

3. The Random Sample

This may be the most important type of sample. A random sample allows a known probability that each elementary unit will be chosen. For this reason, it is sometimes referred to as a probability sample. This is the type of sampling that is used in lotteries and raffles. For example, if you want to select 10 players randomly from a population of 100, you can write their names, fold them up, mix them thoroughly then pick ten. In this case, every name had any equal chance of being picked. Random numbers can also be used (see Lapin page 81).

Types of Random Samples

- a. A **simple random sample** is obtained by choosing elementary units in such a way that each unit in the population has an equal chance of being selected. A simple random sample is free from sampling bias. However, using a random number table to choose the elementary units can be cumbersome. If the sample is to be collected by a person

untrained in statistics, then instructions may be misinterpreted and selections may be made improperly. Instead of using a least of random numbers, data collection can be simplified by selecting say every 10th or 100th unit after the first unit has been chosen randomly as discussed below. such a procedure is called systematic random sampling.

- b. A **systematic random sample** is obtained by selecting one unit on a random basis and choosing additional elementary units at evenly spaced intervals until the desired number of units is obtained. For example, there are 100 students in your class. You want a sample of 20 from these 100 and you have their names listed on a piece of paper may be in an alphabetical order. If you choose to use systematic random sampling, divide 100 by 20, you will get 5. Randomly select any number between 1 and five. Suppose the number you have picked is 4, that will be your starting number. So student number 4 has been selected. From there you will select every 5th name until you reach the last one, number one hundred. You will end up with 20 selected students.
- c. A **stratified sample** is obtained by independently selecting a separate simple random sample from each population stratum. A population can be divided into different groups may be based on some characteristic or variable like income of education. Like any body with ten years of education will be in group A, between 10 and 20 group B and between 20 and 30 group C. These groups are referred to as strata. You can then randomly select from each stratum a given number of units which may be based on proportion like if group A has 100 persons while group B has 50, and C has 30 you may decide you will take 10% of each. So you end up with 10 from group A, 5 from group B and 3 from group C.
- d. A **cluster sample** is obtained by selecting clusters from the population on the basis of simple random sampling. The sample comprises a census of each random cluster selected. For example, a cluster may be some thing like a hospital, a village or a school, a state. So you decide all the hospitals in a certain geographic area are clusters. You want 20 hospitals selected. You can use simple or systematic random sampling to select the hospitals, and then every hospital selected becomes a cluster. If you interest is to interview doctors on their opinion of some new program which has been introduced, then all the doctors in a cluster must be interviewed. Though very economical cluster sampling is very susceptible to sampling bias. Like for the above case, you are likely to get similar responses from teachers in one school due to the fact that they interact with one another.

Purposeful Sampling

Purposeful sampling cases depend on the study purpose. There are about 16 different types of purposeful sampling. They are briefly described below for you to be aware of them.

Extreme and deviant case sampling this involves learning from highly unusual manifestations of the phenomenon of interest, such as outstanding successes, notable failures, top of the class, dropouts, exotic events, crises.

Intensity sampling this is information rich cases that manifest the phenomenon intensely, but not extremely, such as good students, poor students, above average/below average.

Maximum variation sampling involves purposefully picking a wide range of variation on dimensions of

interest. This documents unique or diverse variations that have emerged in adapting to different conditions. It also identifies important common patterns that cut across variations.

Homogeneous sampling reduces variation, simplifies analysis, and facilitates group interviewing. Like instead of having the maximum number of nationalities as in the above case of maximum variation, it may focus on one nationality say Americans only.

Typical case sampling involves taking a sample of what one would call typical, normal or average for a particular phenomenon.

Stratified purposeful sampling illustrates characteristics of particular subgroups of interest and facilitates comparisons between the different groups.

Critical case sampling permits logical generalization and maximum application of information to other cases like "If it is true for this one case, it is likely to be true of all other cases. You must have heard statements like if it happened to so and so then it can happen to anybody. Or if so and so passed that exam, then anybody can pass.

Snowball or chain sampling identifies cases of interest from people who know people who know what cases are information rich that is good examples for study, good interview subjects. This is commonly used in studies that may be looking at issues like the homeless households. What you do is to get hold of one and he/she will tell you where the others are or can be found. When you find those others they will tell you where you can get more others and the chain continues.

Criterion sampling has a set of criteria and all cases that meet specified criteria are picked. For example, all ladies six feet tall, all white cars, all farmers that have planted onions. This method of sampling is very strong in quality assurance.

Theory based or operational construct sampling. Finding manifestations of a theoretical construct of interest so as to elaborate and examine the construct.

Confirming and disconfirming cases are used to elaborate and deepen initial analysis such as if you had already started some study, you are seeking further information or confirming some emerging issues which are not clear, seeking exceptions and testing variation.

Opportunistic sampling involves following new leads during field work, taking advantage of the unexpected flexibility.

Random purposeful sampling adds credibility when the purposeful sample is larger than one can handle. Reduces judgment within a purposeful category. But it is not for generalizations or representativeness.

Sampling politically important cases attracts or avoids attracting attention undesired attention by purposively eliminating from the sample political cases. These may be individuals, or localities.

Convenience sampling is used to get general ideas about the phenomenon of interest. For example you decide you will interview the first ten people you meet tomorrow morning. It saves time, money and effort. It is the poorest way of getting samples, has the lowest credibility and yields information-poor cases.

Combination or mixed purposeful sampling combines various sampling strategies to achieve the desired sample. This helps in triangulation, allows for flexibility, and meets multiple interests and needs. When selecting a sampling strategy it is necessary that it fits the purpose of the study, the resources available, the question being asked and the constraints being faced. This holds true for sampling strategy as well as sample size.

Sample Size

Before deciding how large a sample should be, you have to define your study population. For example, all children below age three in a certain geographic area. Then determine your sampling frame which could be a list of all the children below three.

The question of how large a sample should be is a difficult one. Sample size can be determined by various constraints. For example, the available funding may determine the sample size. When research costs are fixed, a useful rule of thumb is to spend about one half of the total amount for data collection and the other half for data analysis. This constraint influences the sample size as well as sample design and data collection procedures.

In general, sample size depends on the nature of the analysis to be performed, the desired precision of the estimates one wishes to achieve, the kind and number of comparisons that will be made, the number of variables that have to be examined simultaneously and how heterogeneous a universe is sampled. For example, if the key analysis of a randomized experiment consists of computing averages for experimentals and controls in a project and comparing differences, then a sample under 100 might be adequate, assuming that other statistical assumptions hold.

In non-experimental research, most often, relevant variables have to be controlled statistically because groups differ by factors other than chance.

More technical considerations suggest that the required sample size is a function of the precision of the estimates one wishes to achieve, the variability or variance, one expects to find in the population and the statistical level of confidence one wishes to use. The sample size N required to estimate a population mean (average) with a given level of precision is:

The square root of $N = (1.96 * (\sigma) / \text{precision})^2$ Where σ is the population standard deviation of the for the variable whose mean one is interested in estimating. Precision refers to width of the interval one is willing to tolerate and 1.96 reflects the confidence level.

For example, to estimate mean earnings in a population with an accuracy of \$100 per year, using a 95% confidence interval and assuming that the standard deviation of earnings in the population is \$1600.0, the required sample size is $983 : [(1.96)(1600/100)]^2$ squared.

Deciding on a sample size for qualitative inquiry can be even more difficult than quantitative because there are no definite rules to be followed. It will depend on what you want to know, the purpose of the inquiry, what is at stake, what will be useful, what will have credibility and what can be done with available time and resources. With fixed resources which is always the case, you can choose to study one specific phenomenon in depth with a smaller sample size or a bigger sample size when seeking breadth. In purposeful sampling, the sample should be judged on the basis of the purpose and rationale for each study and the sampling strategy used to achieve the studies purpose. The validity, meaningfulness, and insights

generated from qualitative inquiry have more to do with the information-richness of the cases selected and the observational/analytical capabilities of the researcher than with sample size.

Conclusion

In conclusion, it can be said that using a sample in research saves mainly on money and time, if a suitable sampling strategy is used, appropriate sample size selected and necessary precautions taken to reduce on sampling and measurement errors, then a sample should yield valid and reliable information. Details on sampling can be obtained from the references included below and many other books on statistics or qualitative research which can be found in libraries.

DAY 7 (PM): ANALYZING M&E DATA

Developing the data analysis strategy is an important part of the planning process. It helps to know the options for data analysis, with the various strengths and weaknesses, as you plan your research. Of course, data analysis is important in the "doing". This overview will provide you with the big picture issues.

Qualitative analysis is best used in situations where we need a fairly in-depth understanding of the intervention, when we are evaluating something relatively new can be used to answer questions like:

- Is the intervention being implemented according to plan?
- What are some of the difficulties faced by staff?
- Why did some participants drop out early?
- What is the experience like for participants?
- Is there any unexpected impact on families and communities?

Quantitative analysis can be used to answer questions like:

- What is the percent distribution?
- What's the average?
- How do participants rate the usefulness and relevance of the intervention?
- How much variability is there in the data?
- What's the relationship between a program and the outcome measures?
- How strong is the relationship?
- Are the results statistically significant?

Analyzing Qualitative Data

Qualitative data analysis is used for any non-numerical data collected as part of the evaluation. Unstructured observations, open-ended interviews, analysis of written documents, and focus groups transcripts all require the use of qualitative techniques. Analyzing qualitative data is challenging, although many people find it interesting. Great care has to be taken in accurately capturing and interpreting qualitative data.

Making Good Notes

When you collect qualitative data, you want to capture as much information as possible. It is important to accurately capture your observations; good notes are essential. This means paying close attention to language: what people say and how they say it. Try not to interpret what they say as you write your notes. Write down anything you observed, any body language, or anything that happened while you were collecting data (for example, many interruptions during the interview). You may also want to capture your immediate thoughts, reactions and interpretations. Keep them in a separate section of your notes. It is extremely important to leave time soon after an interview, observation, or focus group to review your preliminary notes and make additions, and write up your notes so they will make sense when you look at them later on. It is surprising how difficult it is to make sense of notes taken in an interview, focus group, or observation session – even from just a few days earlier. Even if you have tape-recorded the session, a small amount of time invested in a preliminary write-up while it is fresh in your mind will save hours and hours of listening to or watching tapes or poring over transcripts later on.

Drawing Out Themes and Patterns

When analyzing qualitative data, your goal is to summarize what you have seen or heard in terms of common words, phrases, themes or patterns. As you review the material, you will begin to make notes. It helps to read a few to get a sense of what is there and develop a general framework for analyzing the rest of the data. However, you may discover other themes as you go along and may have to re-read earlier material. As you identify the words, issues, themes or patterns, identify where they are located so you can find them again if you need to verify exact quotes or context. This will be very tedious the first time you do it; as you gain experience you will find you can locate potentially important information much more quickly.

Some people find it helpful to use a spreadsheet that identifies the common themes and where they are located in their notes. Other people use note cards to sort through the qualitative data.

DAY 8 (AM): ANALYZING M&E QUALITATIVE DATA

Content Analysis

Content analysis is a systematic approach to qualitative data analysis that identifies and summarizes message content. We usually use the term content analysis to refer to the analysis of such things as books, brochures, written documents, speeches, transcripts, news reports, and visual media. A typical content analysis might be to examine the content of children's textbooks to see whether they cover the necessary material for learning a particular subject, presented in a way that is reading level appropriate and fits the context in which the children live and study. A deeper analysis might examine whether the textbooks convey a specific political agenda or biased interpretation of history.

Sometimes content analysis is used to when working with narratives such as diaries or journals, or to analyze qualitative responses to open-ended questions on surveys, interviews or focus groups.

Summarizing Qualitative Data

Generally, you will report your qualitative data in terms of "common themes" or "a number of people said..." Sometimes there is an isolated idea or perspective that you may want to highlight even though it is not a common theme. There is no rulebook here since the analysis is very much embedded in the particular situation.

Sometimes it is useful to count the incidence of specific themes to give some sense of how prevalent a particular line of thought is among respondents. For example, you might want to specify that X% of the news stories had a liberal bias as compared to Y% that had a conservative bias. If you are using a mixed method (qualitative and quantitative) data collection approach, you will want to find themes and comments that help clarify and illuminate some of the qualitative data. For example, if 55% of the respondents were dissatisfied with the accessibility of the intervention, it helps to have a representative mix of comments that help illustrate what kinds of problems people were experiencing.

You will want to capture some of the "quotable quotes." These are the actual statements of the participants and are chosen because they clearly present a theme or an important point you want to emphasize. There is power with these words, so select them carefully. You may find that your audience will be more likely to remember a quote than a page of description. Be careful that you do not introduce bias here. You may want to present several different quotes that show the range of issues and perspectives about the same theme.

One example of an evaluation that used primarily qualitative methodology was an experiment to decentralize schools. The evaluation collected the views of teachers, parents, and school directors, and one of the themes that emerged was "autonomy."

Controlling for Bias

There is some risk of bias in working with qualitative data in particular; we often see what we want to see and genuinely miss things that don't conform to our expectations. It helps (but does not always completely remedy) to have another person analyze the data. By comparing the two analyses, new themes or different ways of understanding the data may emerge. When reporting qualitative data, you will not be able to

present a count of how many or what percent said or did something. Since all were not asked the same question, you really don't know how everyone felt about that question.

When conducting a content analysis, for example, evaluators review documents and code them in terms of themes. The coders must be trained. Having two people read and code the same set of documents helps better control for individual differences in perceptions. If your evaluators are well trained and the operational definitions and rating systems are clear and agreed upon in advance, both evaluators would have a high rate of agreement in their ratings of the material. The jargon term is “inter-rater reliability” and a high rate would be an indicator of credibility.

Affinity Diagrams

If you are working with several people on the project, you may find it helpful to have each person identify what they believe are the common themes or interesting points they would like to report. An affinity diagram is a good strategy to use here. A good process for this is to have people put one idea or theme on a file card or post-it note, and have everyone place all their cards or post-its on a wall. As a group, you can then sort them into similar ideas and themes. By using this process, everyone's ideas are considered and there is less ownership of a single idea. It is also a very quick way to develop an organizing structure for the analysis and final report.

Concluding Thoughts on Qualitative Data Analysis

Many of us have spent much of our lives being somewhat phobic about statistics. As a consequence, there is a strong tendency to think that the use of qualitative methods is somehow the easy option. But as we have seen in this section, there is a lot more to doing good qualitative data analysis than meets the eye of the casual observer. Analyzing qualitative data is very labor intensive and time consuming, but can reveal some of your most valuable information. Be sure to plan enough time to do this well. Qualitative methods can be powerful tools for looking at causality – whether observed changes are due to the intervention or to something else.

DAY 8 (PM) ANALYZING M&E QUANTITATIVE DATA

Quantitative data are analyzed using statistics. This section will introduce you to some of the most important statistical concepts you need to know as a user and conductor of development evaluations.

Statisticians divide statistics into two large categories. One is **descriptive statistics**, which in its narrowest definition is used with census or nonrandom sample data. The second category is **inferential statistics**, which is used with random sample data. While there are some data analysis techniques that are used only with inferential statistics, many can be used with both kinds of data. This overview will start with the most common data analysis techniques used for descriptive data, and then it will focus on commonly used data analysis techniques for data obtained using random samples.

Descriptive Statistics

Distributions describe the frequency and/or percentage distribution of a single variable. It tells you how many and what percent: 33% of the respondents are male and 67% are female.

Analyzing Quantitative Survey Results

Non-open-ended survey results can be reported in terms of percent answering (e.g., 52% women, 48% men). Sometimes, the questions ask for specific counts (e.g., “Were you employed in the past week?” or “How many goats do you own?” These would also be reported in terms of percents or averages. Other times, people are asked to give opinions along a scale. For example, one may ask whether the respondents have been able to apply what they have learned, and are given a scale ranging from “Not at all” to “A lot.” This is a five-point scale. An example is: 1) agree strongly, 2) agree, 3) no opinion, 4) disagree, 5) disagree strongly. When analyzing this type of data, you would want to establish a decision rule: will you focus on the percent who answered at the extreme ends of the scale (1 or 5), or will you focus on those who answered on either side of the middle category (2 or 4)? Some guidelines might be helpful but there are no firm rules here.

Guidelines for Analyzing Quantitative Survey Data

1. Choose a standard way to analyze the data and apply it consistently.
2. Focus on one side of the scale or the other.
3. Do not combine the middle category with each side of the scale.
4. When reporting on one category out of the 5, choose an extreme category (strongly agree or strongly disagree).
5. Do not report an agree or disagree category without also reporting the strongly agree or strongly disagree category.
6. Analyze and report percentages (not numbers, and definitely not averages).
7. Provide the number of respondents for an anchor.
8. If there is a little difference in the data, raise the benchmark: what do the results look like when you focus on the questions that received a majority saying “very satisfied”?
9. It is an art and a skill; it gets easier with training and practice.

Describing Two Variables at the Same Time:

Sometimes you want to describe two variables at the same time. For example, suppose you want to describe the composition of the hands-on and lecture classes. For each class, you want to know what percent were boys and what percent were girls. Analysis of the data shows that the hands-on classes were comprised of 55% boys and 45% girls, while the traditional lecture classes were comprised of 55% girls and 45% boys.

	Girls (%)	Boys (%)	(%)
Hands on	45	55	100
Traditional	55	45	100
	100	100	

Measures of Relationship:

Measures of relationship (or association) tell you how strongly variables are related. While they do not prove “cause,” you cannot even begin to suggest a causal argument unless the variables are strongly related (or associated). Causation requires showing that 1) the two variables vary together, are associated, 2) one variable precedes the other variable in time and 3) no other explanations for the observed results are plausible. In other words, a strong measure of association is required but it is insufficient by itself to prove causation.

While there are many kinds of measures of association, they are all reported either in terms of 0 to 1 scale or -1 to + 1 to indicate the strength of the relationship. If it were a perfect relationship, it would get a score of 1. If there were no relationship, it would get a score of 0. The closer the measure is to 0, the weaker the relationship. The closer the measure is to 1, the stronger the relationship.

Some measures of association are also calculated to show the direction of the relationship. It shows that through the sign (positive or negative). A measure with a positive sign means that as the variables change in the same direction: both go up or both go down. This is called a direct relationship, increases. A negative sign says that the variables have an inverse relationship, meaning, they move in the opposite directions. For example, as age increases, health decrease. A measure of association of -1 would therefore mean a perfect inverse relationship. A measure of association of -.1 would be close to zero and is a very weak inverse relationship.

Correlation is another measure of association. You can do a correlation between two interval or ratio variables or use a multiple regression technique to estimate the impact of several variables simultaneously on the dependent variable. This technique works with interval and ratio level data. There are other techniques that can be used for ordinal and nominal data but they are less commonly used and are harder for the average person to understand.

Inferential statistics enable you to make an estimate about a population based on a random sample selected from that population. Whenever you are using sample data, your major concern is whether the results are a function of some quirkiness of your sample rather than a true picture of the population. If you had picked a different sample, would your results be fairly similar?

Statisticians have developed tests to estimate this. These are called statistical significance tests and do a very simple thing: They allow you to estimate how likely it is that you have gotten the results you see in your analysis by chance alone. Statistical tests come in 100+ varieties. You may have heard of some of the more common statistical tests, such as Chi Square and the t-test. The good news is that all the different statistical tests are interpreted using the same guidelines. Evaluators typically set the benchmark for statistical significance at the .05 level. This is sometimes called the alpha level or the p-value. That is, we set the benchmark so that we are at least 95% certain that the sample results are not the result of random chance. If we want to raise the bar, we would set the level at .01 to be 99% certain that the sample results are not due to chance alone.

All tests of statistical significance are partly based on sample size. If the sample is very large, small differences are likely to be statistically significant. You still need to decide whether the differences are important given your research. Importance is always a judgment call. Just because a result is statistically significant does mean it is important or meaningful.

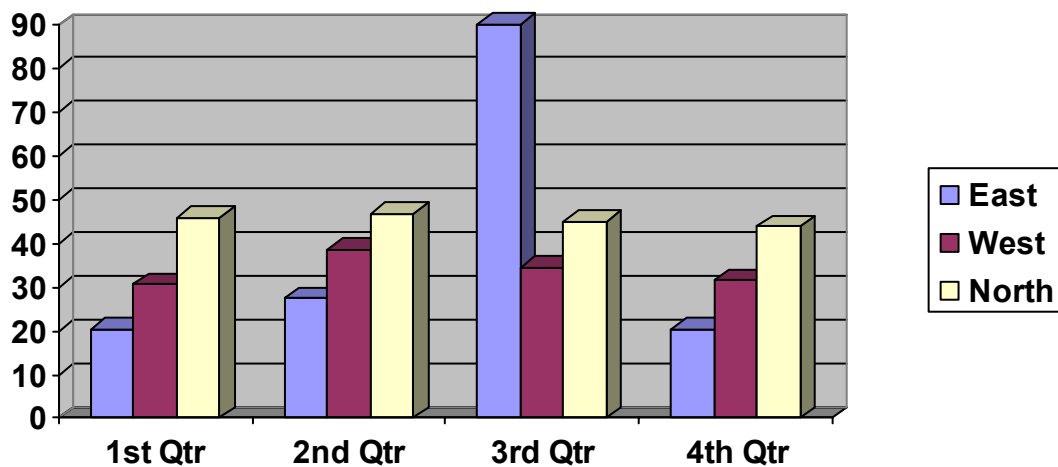
For example, suppose results from a random sample of people show that 85% are satisfied with the care they receive from the public health clinic and 89% are satisfied with the care from private physicians. These results are statistically significant at the .05 level. So what? While there is a slight difference, most people are satisfied with both types of care. While this difference is not likely to be the result of chance alone, it is not a result that would be used as a basis for making change to policies about health care providers.

The Logic of Statistical Significance Testing

The tests are set up to give you the probability of getting the results you got if there really was no difference in the population as a whole. Researchers present it in terms of the “null hypothesis.” The null hypothesis is always one of no difference. Suppose a survey based on a random sample of people in Pakistan shows that there was a 5,000 rupee difference in annual income between men and women. It might be framed this way: if there really is no difference in the population, what is the probability of finding a 5,000 rupee difference in income between the men and women in a random sample? If there is a 5% chance (.05) or less (that's our benchmark), then we will conclude that the sample results are an accurate estimate of the population.⁴ We would conclude that there is indeed a difference of about 5,000 rupees, and that difference is **statistically significant**. Technically, social scientists would say that the null hypothesis of no difference is rejected. If, on the other hand, the test for statistical significance is .1 (which is greater than .05), social scientists would say that the test failed to reject the null hypothesis: these results are not statistically significant.

Analyzing Trend Data

Sometimes the analyst has data that has been collected over time. This data series can be analyzed to see if there is a dramatic shift in the data series – upwards or downwards- over time. The table below shows the relative shifts of three trend lines over four quarters of data. This type of trend analysis – using three data series – can show the relationship among three different variables and if it changes over time



The main requirement of the trend analysis is that the observations representing the data at each time period in the trend chart – e.g., over the four quarters – were obtained in the same manner over time. That is, the data are equivalent over time in terms of how measurements were made and the data collected. If these conditions can be met, then trend analysis offers an informative means of assessing changes over time and the relationships among several variables. The Excel spreadsheet platform offers a useful tool for both archiving and analyzing the M&E data.

DAY 9 AM: COST-EFFECTIVENESS ANALYSIS

Cost-effectiveness analysis involves coupling the intervention cost per client with effectiveness measures (both described above). Various outcomes can be calculated, depending on the availability of data. For example, recent cost-effectiveness analyses have been conducted of the impact of nevirapine on HIV transmission from mother to child, and on the impact of enhanced STI services on reducing HIV transmission. Below is a summary of the most commonly used outcomes:

- Cost per unit of behavior or change—This includes outcomes such as the cost per condom used or the cost per partner reduced. This approach is not very common in the literature.
- Cost per HIV infection averted—This is a frequently used approach.
- Cost per QALY or DALY saved—Society may place a greater value on averting HIV cases among persons with certain characteristics (especially with regard to the quality of life they will experience) and among persons who will have a longer life should infection be averted. Thus, techniques exist to weight the impact of the infection averted and take into account the quality of life and the number of years of saved life that will result from an intervention. The two most common approaches are to convert the HIV infection averted to Quality Adjusted Life Years (QALY) or Disability Adjusted Life Years (DALY). To convert the number of HIV infections averted into QALY or DALY requires that the age of infection of the target population be identified. Discrete stages of infection are then identified, each with an associated time duration and weight for the quality of life. The weighted number of years of life saved from the intervention are calculated using the age specific distribution of the target population. There is a growing consensus in the field on how to do this for HIV. However, most applications have been based on U.S. data and information on the natural history of HIV in developing countries is not well understood due, in particular, to the lack of long-term cohort studies that include quality of life measures.
- Cost utility analysis—This approach uses cost-effectiveness measures described above (mostly QALY and DALY) and takes into account the treatment costs of HIV at different stages. Simple formulas can be used to calculate the cost-utility ratio, which is a common measure used in health economics¹⁰. The benefit of the cost-utility ratio is that it places the cost-effectiveness of the intervention in relation to the cost-saving from treatment that result. This allows for direct comparisons to other health interventions.

Assessing the Analysis Strategy

This assessment tool can be used to determine the type of analysis that is possible with different types of data. In some cases, there are options for the outcome variable that is examined. The minimum level of data required is highlighted with asterisks. As described above, four primary types of analysis are possible: (1) cost analysis, (2) effectiveness analysis, (3) modeling of effectiveness, and (4) cost-effectiveness analysis. Cost-effectiveness analysis can be broken out into four additional areas, including: (1) cost per unit of behavior change, (2) cost per HIV infection averted, (3) cost per QALY or DALY saved, and (4) cost-utility analysis, which incorporates the treatment costs averted from the intervention. The next section looks more closely at cost-effectiveness analysis and reviews the steps involved in estimating the cost-effectiveness of an HIV intervention.

Guidelines For Conducting A Cost-Effectiveness Analysis

In this section, the steps normally taken to conduct a cost-effectiveness analysis are reviewed.

1. *Framing the Problem*

It is crucially important to frame the problem carefully before initiating a cost-effectiveness analysis. This involves specifying the study question, which in turn helps to define the other key elements of the analysis: perspective, time frame, and analytic horizon. Example of study questions include:

- What is the most cost-effective behavioral intervention for in-school sexually active adolescents in Western Kenya?
- Is it more cost-effective to target older or younger adolescents in an HIV peer education intervention?
- Are peer education programs more cost-effective than HIV voluntary counseling and testing among factory workers in Thailand?

The perspective of the analysis relates to the question, "Who is responsible for the costs and consequences of the program being evaluated?" For example, analysts may be interested in examining the perspective of the donor (such as USAID or the national government) who pays for the intervention.

Other perspectives that might be examined include the entire society, the government, the implementing agency, and the individual who receives the intervention. The choice of perspective helps to determine the types of costs that are captured. For example, a donor may incur the costs of the intervention, but not the costs of treatment for HIV infections in a given setting.

The time frame for the analysis should then be determined. In evaluating HIV interventions, a 1-year time frame is typically used. Thus, costs are generated that reflect the annual cost per client, and the annual number of HIV infections averted. It may be useful to also expand the time frame to include more long-term impacts, although it is often difficult to access data on long-term impacts of HIV interventions. In situations in which the time frame is a number of years, it is necessary to discount the costs to the present year value. There are techniques and formulas for discounting, and one simple approach is to use the following formula: $(1 + r)^{-t}$, where "r" is the discount rate (usually set to 3% to 5%), and "t" is the number of years from the current year.

The analytic horizon should next be determined. This is the amount of time over which the outcome is examined. For example, an HIV intervention may operate for only 2 years, but the benefits of the intervention may be realized over a lifetime for the recipients of the intervention. Thus, the analytic timeframe should probably be 3 years in this instance. For HIV interventions, care should be taken to assure that adequate data are available regarding outcomes over time. In most cases, it is probably better to set the time horizon at a shorter interval to generate a more conservative analysis. If long-term estimates are made, the results may have poor validity, and will be difficult to defend and use for policy purposes.

2. *Identifying the Options to be Compared*

Once the study question has been determined, it is important that the comparisons for the analysis be carefully defined. In many cases, an HIV intervention will be compared to no intervention. Alternatives include comparisons between two different interventions, or comparisons between two or more target populations. Comparisons can also be made between the health benefits of an HIV intervention versus a non-HIV health intervention. Various references are available to assist in making such comparisons¹¹. It is important that data for the analysis defined at this step be available.

The more complex the comparisons defined, the more difficult it will be to collect the requisite data for the analysis. Additionally, for purposes of analysis, it is typically necessary to have, at a minimum, pre-intervention and post-intervention data on the effectiveness of the intervention. That means that if two intervention approaches are being compared, four data sets are needed—data for before and after the intervention for each intervention approach.

3. *Identifying the Outcome Measures*

This is perhaps one of the most difficult steps in a cost-effectiveness analysis for HIV interventions. Again, some of the most frequently used outcome measures for HIV intervention cost-effectiveness analysis include the cost per HIV infection averted, the cost per quality adjusted life year saved, and the cost-utility ratio. The cost to the client may also be of interest in these analyses. Additionally, it may be useful to examine medical and social costs and benefits of the intervention, although these are often difficult to measure.

4. *Identifying Intervention and Outcome Costs*

Next, the costs of the intervention and outcomes need to be estimated. Care should be taken to harmonize the costs estimates with the comparisons made, and the analytic horizon of the analysis. With HIV intervention studies, the costs of the intervention are typically conveyed in terms of the cost per client to receive the intervention.

5. *Conducting a Sensitivity Analysis*

The final step in the cost-effectiveness analysis is to conduct a sensitivity analysis of the decision tree and its associated analysis models in situations where outcomes are modeled. Sensitivity analysis takes into consideration any uncertainty that occurs in the data used. In all scientific studies there is some level of uncertainty in data that are collected and used for analysis (the speed of an atom, the size of a tumor, the chance of a volcano eruption). One way to capture and analyze the effect of such uncertainty in data is to first model the system mathematically, and to then make systematic changes in the parameters used in the mathematical analysis to see how they affect the outcome. By varying uncertain values over a reasonable range, it is possible to examine changes in variables in the system and see how stable the system is when values are changed. Selecting a reasonable range of values to represent the uncertainty is an important qualitative process that must be conducted carefully.

An example of an important sensitivity analysis that needs to be conducted with each intervention is the likely range of effectiveness that will be generated from the intervention. One good possibility is to use the confidence intervals of the outcome of interest, such as HIV incidence, to set the high and low values for the sensitivity analysis. In situations where one models the outcome, confidence intervals can be used for the input parameters, such as with percent condom use with the AVERT model. The process for conducting

sensitivity analysis is to run the analysis multiple times with varying parameter values. This simple form of sensitivity analysis (known as a one-way sensitivity) examines how changes in individual variables affect the outcome of interest. More sophisticated sensitivity analyses can be conducted that examine how sets of input parameters act together to affect the outcome. These typically are conducted with specialized software, such as At-Risk6.

6. *Presenting the Results*

Once all of the requisite analysis is completed, it is important to develop a policy presentation of the results. In doing this, it is important to consider the audience for the information. Overly technical presentations to persons not familiar with cost analysis can result in a poor response. It is also helpful to give concrete examples and to present the following details.

- the study question;
- the study perspective, time frame, and analytic perspective;
- the assumptions used to build the model and estimate outcomes;
- a description of the interventions;
- evidence of the effectiveness of the
- interventions;
- identification of the relevant costs,
- including whether productivity costs are included, and the discount rate used;
- results of the analysis showing the comparisons made;
- results of the sensitivity analysis;
- discussion of the results that incorporate the social and policy perspective; and
- recommendations for action.

Conclusion

Cost-effectiveness analysis can provide important insights into the utility of HIV intervention programs. Care needs to be taken to select an analytic approach that matches the specific research question of interest and available data. When conducting a cost-effectiveness analysis of an HIV intervention, it is important to carefully develop the analysis plan. The approaches outlined here describe basic methods for conducting a cost-effectiveness analysis. More sophisticated techniques not covered in this session are also available. It is recommended that use of advanced techniques be done cautiously and with the technical assistance of persons with experience in this area.

Exercise: Participants discuss the potential for using cost-effectiveness in their programs. Do they see and barriers to its use? How could they be dealt with? Can they think of a specific example of where it could be used?

DAY 9: (PM) USING THE M&E RESULTS

M&E is empowering. By participating in a M&E exercise, you have had a hand in shaping the information through which someone can come to understand your program's purposes and accomplishments. You have also provided yourselves with a powerful tool for improving and expanding your program and its activities in fruitful ways. And you are very likely also to be in a position to influence the further development of supports for women victims of violence in your community. The benefits provided by a good evaluation can make all the effort seem worthwhile.

Once you have gone to the trouble to participate in evaluation activities, how can you make sure that you receive the benefits. The first, and obvious, way to use evaluation information to show others what you have accomplished—that is, you would use the information to promote your program, and to avoid having your program's purposes and achievements misunderstood by various audiences. The second way to use evaluation information is to improve your program. The third way to use it is as a lever for stimulating your community to make changes.

Uses for Program Promotion

A good evaluation should help you promote your program. It should make your goals clear, and should produce information to show how you are moving toward achieving them. You can include this information in news stories to make sure your program is properly understood, use the information in grant proposals to get more money for your program, or use it to tell a variety of audiences about what you do, that you do it well, and that they would benefit from using your services if they ever need them.

Avoiding Misunderstandings of Your Program

When you are responsible for designing your own evaluation, you will rarely be in a situation where you feel the data collected do not adequately reflect your program's goals or accomplishments. But what if you are in the situation of having an evaluation imposed on you by a funder or by a higher level unit in your own organization, and you feel that the data requested will not tell the story of your program? What can you do then?

Suppose, for example, that you run a counseling service and the only thing the formal evaluation wants to know about you is who your clients were. They are asking for the usual demographic information (age,

race, number of children); but they also are asking for some types of data that you think reflect an underlying suspicion of women who are victims of violence (e.g., chemical dependency of the victim, number of times victimized, number of prior relationships in which victimized [if a battering case], or prior sexual behavior [if a sexual assault case]). Further, they are NOT asking anything about what the women have done or are doing to help themselves. They are NOT asking for any documentation of the array of services you have given to the women. Nor are they asking for any evidence that the women appreciated your service, changed their behavior because of your service, felt better because of your service, understood their rights better because of your service, understood their feelings better because of your service, or any other impact of your service.

What should you do? There is no EASY solution to this problem, but that does not mean there is NO solution. What you would need to do is to collect, on your own, the information you feel DOES properly reflect your program's goals. Once you have this information, be sure to present it alongside the formal evaluation every time those formal data get reported. Develop materials that show what your program really does, and how women really feel about it. Presenting this information augments any impressions created by the minimal victim profile data with the broader picture of your clients and their involvement with your services.

There are also several last steps you could try. First, even before you collect this additional data, you could try working with other programs in your position to develop a data collection protocol that you commit to use *in common*. Then there will be strength in your numbers, both methodologically (that is, you will have more data), and politically (that is, you will have more programs working together). Second, once you have your additional data, you should send the results to whomever imposed the oversimplified evaluation. Ask that they be examined and incorporated in any reports that are issued describing evaluation results. Finally, you could try to reopen negotiations with the overall evaluation/evaluator/funder to see if in the future, the evaluation can be expanded to include more appropriate indicators of program activities and accomplishments. This might be particularly appealing since you will have demonstrated that you can provide the relevant data.

Using Evaluation Information to Get Funding

To use evaluation information to support fund-raising efforts, you would (1) identify the primary mission of any funding source you want to approach; (2) select the data from your evaluation that shows that you are addressing this mission and doing so successfully; and (3) using these data, gear your presentation to the funding source's interests and purpose.

Suppose, however, that you think there are some things that have to be accomplished before the funding source's goals can be accomplished. Your program does some of these things, and you want a funding source to support your activities, even though you can't promise that you will reach their goals every time, or quickly. This is where your logic model can stand you in good stead, assuming that you have thought it out well and that your evaluation has documented all the intermediate steps between what you do and some ultimate goals.

Take training as an example of a preliminary step, and "more convictions" as the ultimate goal. You do training for health workers, and cannot in any way guarantee positive outcomes. But if you have done your logic model well, you will be able to document a number of steps that should, ultimately, produce better results. For instance, you could be able to show (if you collect the right data) that compared to those who did not receive training, health workers receiving training are more thorough and careful in their patient interviewing.

Suppose, however, that after several years of training the patient records show that there have not, in fact, been increases in information quality and your funder wants to end its support for your activities. You can use evaluation techniques to learn where the problem lies. Have workers highlight those records prepared post-training to show directly the impact of the training on record keeping and pinpoint the exact ways in which the process can be improved and the steps to address poor practices.

Using Evaluation Information to Increase Public Appreciation of Your Program

Program evaluation data can also be used in all sorts of promotional ways. You can include them in brochures, flyers, and annual reports that you distribute all over town. You can hold an "Open House" to which you invite reporters, representatives of other agencies that have contact with women victims of violence, representatives of agencies and organizations concerned in general with women's well-being, and so on. Create attractive posters showing your program in action, and include on each poster one or two sentences or statistics that tell what you have accomplished. You can develop good relations with local reporters, if you don't already have them. When they write stories about your program, be sure the stories include one or two "boxes" or "sidebars" that use evaluation data to show your successes.

Uses for Program Development

Perhaps the most important use for evaluation results is program improvement and development. Evaluation data are particularly useful for helping you look at your program and see what is going wonderfully, what is okay but could be improved, and what cries out for fixing (most programs have some of each). Think about evaluation as an ongoing opportunity for reflection, and for comparing your performance against what you hope your program will achieve.

From the beginning, getting ready for evaluation helps you think about your program and how it is organized. You must be able to sit down and describe to an outsider what you are trying to accomplish (your goals), how the activities you perform every day increase the chances that you will reach your goals, and what internal and external factors might make it easier or more difficult for you to reach your goals. You cannot do this well on the spur of the moment. Many programs set up a retreat or off-site meeting to do this, with an evaluator and possibly also with a facilitator. This gives program staff the luxury of sitting back and thinking about how their program works; usually no one has done this for many years, if ever. In doing this exercise, many programs identify strengths of which they are proud. However, they usually also identify weaknesses, areas of misdirected energy, issues of whether current time allocations are the best use of resources for accomplishing their goals, etc. In short, the opportunity for reflection afforded by preparing for an evaluation can stimulate program improvements even before the evaluation formally begins.

A premise on which a good evaluator operates should be "no surprises." You don't want to get to the end of an evaluation, having heard nothing about findings for its duration, and be hit all of a sudden with a thick report. Even if the findings are overwhelmingly good, waiting until the end to learn them gives you very little opportunity to absorb them and figure out what they mean for potential changes in program operations. If there are some negative findings or findings about areas that need improvement, it is a lot more useful to learn about these findings as they emerge so you can discuss them and decide what to do about them. Getting a lot of negatives dumped on you at one time is discouraging and not likely to be productive, in addition to which it does not make people feel good about continuing to do evaluations.

You should set up regular feedback sessions with your evaluator to discuss evolving findings related to program processes and activities, as well as to get the perceptions and feelings of the evaluator as she or he spends time in your program and with your clients. This interaction can help both the program and the evaluator. The program staff can help the evaluator interpret the meaning of emerging findings and offer suggestions for how to gather additional information relevant to developing a full understanding of anything interesting. At the same time, program staff can use the feedback to think about whether the program needs to make some changes in the way it operates, either to enhance good performance or compensate for areas where performance may need to improve.

Another source of feedback on program operations is performance monitoring data. If you have set up data collection on program activities, services, and clients so your data system produces monthly or quarterly reports, you could present them as part of regular staff meetings and invite discussion about what they mean and how your program should respond to them. This type of open and shared discussion can help bring all staff back to an awareness of overall program goals and how daily behavior is or is not contributing to them. In most programs it is all too easy for everyone in a busy program to get overwhelmed with daily coping so they never have this type of discussion.

Regular feedback sessions from the evaluator and from a data system throughout the evaluation are the ideal. However, even if you get the results of evaluation all at once through a report from an outside evaluator, you should still set aside time for the whole staff to review them, absorb their meaning, and make decisions about how the program might want to change in light of the results. A retreat is often a helpful way to accomplish this. The evaluator should be present for part of the time, to help interpret and understand the results, and perhaps might be absent for part of the time while staff discuss what they have heard and what they might want to do about it.

Uses for Community Development

Even in situations where all relevant agencies are "at the table," ongoing evaluation results can be used to improve community coordination and perhaps develop new and more appropriate services. Suppose you have a council, task force, or coordinating body in your community, and it is looking for information that will help to prioritize new projects. Feedback from program clients, systematically collected through a common questionnaire regardless of which agency or service they use, could be one way to pinpoint what needs doing the most. Polls of staff in member agencies about where the system breaks down, what types of help they need from other agencies so they can do their job better, etc., are another source of useful information for establishing priorities. Cross-training sessions are also a way to begin community development. In these sessions, staff of each agency help staff from other agencies learn the agency's primary mission, purposes, ways of functioning, issues with other agencies, and needs for cooperation from other agencies. Everyone has a turn, and everyone stands to gain.

One could even think of developing regular forums for sharing agency news, new programs, findings from research that might interest people in other agencies, and so on. Some communities have monthly lunch meetings attended by as many as 50 or 60 people from every agency in town whose work involves the delivery of health services. Sometimes one agency is "on," to bring the others up to date on things happening in that agency of relevance to health services, sometimes to share data, sometimes to propose new activities. Other presentations may be about a particular problem or issue, such as having someone come to explain the implications of a new law, or deciding that everyone needs to be present at the beginning of a discussion of women whose cases fall through the cracks still left in the system. Once or twice a year, these meetings can be used to present evaluation data and discuss their implications.

Collecting Data over Time

Whatever route you take to using data for community development, having the same (or better) data year after year can make an important contribution. When your community begins its activities, data about current functioning can give you a baseline against which to track your progress. Getting feedback every year (or more often if possible) about how you are doing on major community-wide goals can be a source of renewed commitment, as well as a way to reassess where you are going and what might help you get there faster. In order for this exercise to feel good, it will have to include feedback on intermediate goals so there is some hope that there will be accomplishments to report. Earlier when we talked about logic models, we emphasized the importance of adopting a realistic time frame for achieving your goals, and including many steps along the way so you could track progress and not get discouraged. The same applies to tracking community-wide progress only more so, as it is harder to change a whole community than it is to change one agency.

DAY 10: (AM/PM): PLANNING THE LOGISTICS OF M&E

Necessary Conditions and Capacities

In the appraisal report, you will find an indicative budget for M&E and a description of how M&E should or could be organized. But getting the M&E system working also means thinking of appropriate incentives, ensuring you have the right and enough human capacity at hand, and thinking about ways of storing and sharing information. Discuss appropriate organizational structures for M&E at start-up. This is critical to the success – or failure – of M&E. It is the moment when negotiations need to reach decisions about each of the partners' responsibilities and information requirements. IFAD-supported projects often have one of two basic organizational arrangements for M&E:

1. M&E is coordinated by an M&E coordinator or unit within the project management unit (and supplemented by external M&E contracts) to facilitate management's quick use of information;
2. M&E is carried out by a separate M&E group – for example, subcontracted to a research institute or located within a government department – aiming to provide more objectivity and independent analysis.

Overall, experiences from many projects suggest that M&E is much more effective when those implementing M&E are part of project operations and decision-makers. This can be supplemented in important ways by more external M&E support. And other innovations are emerging that place M&E firmly in the hands of primary stakeholders or field teams.

Questions to guide the detailed planning of M&E conditions and capacities

Conditions and Capacities

Human capacity for M&E

- What are the existing M&E capacities with project partners?
- What training will be necessary?
- What consultancy support will be required?

Incentives for implementing M&E

- Are M&E responsibilities included in job descriptions and terms of reference?
- How will reflection and learning among staff, partners and the intended primary stakeholders be encouraged?

Organized structures

- Will there be an M&E unit or will M&E be spread among all parties? If there is a unit, how many people will it have and where will it be located, under whose authority?
- How closely connected with M&E staff will be with project management? Are the existing M&E capacities with

Management information systems (MIS)

- What information must be stored and accessible, when, how and for whom?
- What level of computerization is required and appropriate?
- What expertise will be required to set up the information management system?

Financial resources

- Is there a separate M&E budget and have sufficient resources been allocated?
- Has the staffing allocation for the project taken into account time for all relevant staff to undertake M&E activities??

Participatory M&E generates attitude and culture of self-evaluation and empowerment in community-based organizations

The participatory evaluation method used by one of the implementing agencies in Prochalate, El Salvador,

The process ends with a final summing-up in which the results of the self-evaluation workshop are compared with the results found by the other team.

Once most of the detailed M&E plan has been completed, you can take a fresh look at the M&E budget. The following example provides a list of likely M&E costs. How you cost M&E depends on whether you allocate resources to specific M&E activities or whether you include M&E in generic categories, such as "staff training", "participatory workshops",

etc.

Possible M&E Costs to Consider in the Budget

Staff time, such as: planning, implementing and improving all the M&E processes; report writing and analysis; capturing and documenting lessons learned; facilitating community-based M&E processes

- Consultants/Technical assistance (fees, travel expenses), such as: developing a detailed M&E plan; establishing management information systems; facilitating review workshops, training and capacity-building; checking of audits
- Evaluation events (venue costs, travel and accommodation, materials, per diems, course fees), such as: M&E planning workshops, annual community review workshops, specific monitoring activities, focused evaluations on important topics
- Materials and equipment, such as: technical equipment for monitoring; computer and network hardware and software; dial-up networking charges; network maintenance contract
- Publication and documentation, such as: printing documents and distribution; display boards; materials

Documenting the Monitoring and Evaluation Plan

Projects have three core documents that serve to guide M&E:

1. The M&E framework in the project appraisal report;
2. The project implementation manual;
3. The M&E plan or manual.

M&E in the Project Implementation Manual

Most projects allocate time and resources to develop what is known as a "project implementation (or procedures) manual". This is a set of guidelines with information about financial accounting procedures, procurement procedures, guidelines for staff travel, guidelines for the use of vehicles and other equipment, and other details necessary for the smooth operation of the project.

The detailed M&E plan may be part of the project implementation manual, an annex to it or a separate document. Irrespective of where it can be found, the implementation guidelines and M&E guidelines must be closely linked and, above all, coordinated. Contradictions or ambiguities in the two sets of guidelines must be avoided.

As the detailed M&E plans may contain an overwhelming degree of detail, summaries for all project participants are helpful to keep everyone focused on their responsibilities. A good way to summarize specific inputs is in an M&E timeline for everyone who plays an important M&E role (see [Box 4-8](#)). Ideally, these timelines should be integrated within weekly and monthly activity timelines so that M&E becomes an integrated part of activities.

Working with an M&E Timeline

Knowing when information is needed is critical. By the time you have reached this point of specifying the M&E plan, you may find it hard to remember all the key M&E moments and how they relate to each other and to planning.

An M&E timeline is a useful tool for maintaining an overview of the various internal activities scheduled, key reporting moments and external missions. One simple aid for coordinating M&E activities is to hang a copy of the timeline in a central meeting room and ask each team member to make a personalized version in which his/her responsibilities are outlined over the year.

Example 1. One project in Zimbabwe records M&E milestones in a calendar like the one below. This project also had a separate monitoring calendar on institutional process that indicated when reports had to be sent to whom and on what topic.

EXTRACT FROM: Calendar on outcome and impact monitoring and assessment (annual)		
Year	Activity	Responsible Person and/or Unit
Y1, 3-4 th quarter	Participatory irrigation scheme appraisal in all schemes	> District team/ planning facilitators
Y2, 1 st and 2 nd quarter	<ul style="list-style-type: none"> Farmer and scheme-level baseline survey (questionnaire) Study on scheme costs Financial viability survey of 24 schemes 	> Project management/ consultant > Project management/ consultant > Planning facilitators
Y2, 3 rd and 4 th quarter	<ul style="list-style-type: none"> Environmental screening/scoping Training needs assessment Service performance assessment of selected rural district councils and support agencies Institutional mapping and SWOT analysis Context assessment 	> Consultant > Training coordinator > Consultant > Consultant > Monitoring expert

Example 2. A project in Nepal plans to use a CEA, a critical events agenda, which lists the most critical project milestones to be monitored. It is to be discussed in annual stakeholder workshops and included in all annual reports. It will be updated regularly. It looks like the table below.

Number	Critical Event		Target Date		Completion Date				Status/ Causes of Delays				
The same project also outlined its indicative monitoring and evaluation plan in the appraisal report in terms of what each M&E report and activity was contributing towards the six main M&E objectives it had set (see below).													
M&E Report and Records	Objectives								Res-pons-ibility	Method-ology	Users		
	Progress Monitoring	Performance Monitoring	Learning	Impact Monitoring	Tracking of Broader Context		Trend Analysis and Forecasting						
Periodic													
19 different entries													
Non-periodic													
8 entries													

Indicative Contents for an M&E Plan

A documented plan is critical for keeping track of activities and resources. The operational plan for project M&E provides the conceptual and, above all, practical basis for planning, monitoring and evaluation within the project. In Cuchumatanes, Guatemala, the M&E document described: its objectives, strategies, methodologies, work plan, its activities in detail and the technical tools to be used. The document also defined the main concepts related to M&E, the redesigned logframe matrix, the adjustment of information collection systems and the database of primary stakeholders.

The M&E operational plan will be the reference point for stakeholders throughout the project life. So it needs to be comprehensive enough, at the macro level, to provide a clear picture of the overall project intentions and how the M&E system will serve this. At the micro level, it must give fine detail on schedules, responsibilities, budgets and so on, which will help guide the drawing-up of AWPBs. Where monitoring tasks are implemented with local stakeholders, such details may have to be translated into local languages.

Indicative Contents for an M&E Operational Plan

A. Purpose and scope

- Project overview and objectives, rationale and justification for the design of the M&E system
- How the M&E system will support project management and meet the reporting requirements and information needs of different stakeholders
- Summary of overall experience of M&E undertaken with key stakeholders
- Discussion of extent of participation, balance between qualitative/quantitative approach, resource intensiveness and the intended poverty focus of the M&E system

B. Approach

Overview of how stakeholders will be involved, what learning-oriented approaches will be used and, in general terms, what information gathering and analyzing methods will be used; for example, the extent of use of participatory approaches, geographic information systems, computer-based information systems or baseline surveys

C. Revised logical framework, plus performance questions, indicators and information needs and sources

Precise definition of all performance questions, indicators and information needs for all levels of the objective hierarchy:

- Assessment of the information needs and interests of all key stakeholders
- Assessment of indicators of exogenous factors and assumptions (e.g., climate, prices, outbreak of pests and disease, economic situation, policy environment)
- Assessment of information needs and indicators for relevance and end-use and for technical and resource feasibility
- Selection of performance questions and indicators

D. Management information and system reporting

- Purpose of the management information system
- Organization of information gathering and synthesis:
- For each expected information product – who, what, when and where
- Schedule of information production – who, what, when, to whom, for what purpose
- How computerized networks and manual archiving systems are expected to function, with/to whom, for which data
- Outline of data storage needs
- Expected reporting outputs, for example:
- Informal communication and feedback channels
- Report flows – deadlines and frequencies

- AWPBs – outline of the AWPB format, including output/activity plans and budgets, consolidated budgets, a training plan, a procurement plan, a contracted services plan
- (Bi-) Annual progress reports for the project as a whole and each component, village-based reviews
- (Bi-) Annual financial reports
- Recurrent supervision missions

E. The M&E work plan (processes and events) and timing of activities

Precise definition of methods to be used with different stakeholder groups for two core purposes:

1. M&E of resources, activities and implementation for effective project operations:
 - Project resources: transport use, allowances, register of assets, register of services/technical assistance
 - Project activities: training (workshops, study tours, etc.), construction (technical or social infrastructure), scheme organization, trials and demonstrations, credit lines, etc.
 - Other monitoring activities
2. M&E of outcomes and impact for guiding the project strategy, for example:
 - Proposed surveys: baseline/household, component, staff
 - Participatory annual assessment and planning workshops
 - Other annual evaluation and beneficiary assessments, reviews and planning sessions
 - Mid-term review and project completion report
 - Feasibility of methods in terms of technology and resources
 - M&E work plan schedule: integrated schedule of key events and reporting/decision-making moments
 - Critical events agenda

F. Establishing conditions and capacities

M&E organization:

- Necessary institutional and stakeholder linkages for M&E
- Existence (or not) of a specific M&E unit and how it relates to the project structure and hierarchy of authority
- Human resource needs:
- Number, capacities and responsibilities of different stakeholders in M&E, including project staff and primary stakeholders
- Incentives for different stakeholders

- Training needs of stakeholders and staff
- Resource needs:
- Vehicles and equipment
- Technical assistance

G. The M&E budget

Detailed budget allocation

H. Appendices

Original and revised logframes

- List of proposed indicators
- Outline formats for data collection, annual and biannual schedule of activities, etc.
- Outline formats for preparing: quarterly, biannual and annual reports; a summary of main project achievements; status reports on project inputs and resources, project outputs and results; evaluation studies – summary of findings and recommendations
- Baseline survey questionnaire
- Staff job descriptions and details of allowances
- Technical Assistance terms of reference
- M&E work plan
- Detailed budget of M&E

Quality of Your Monitoring and Evaluation and Keeping It Updated

Once you have a detailed M&E system, two more steps are needed. First, you need to check the overall quality of the system itself, as designed. Second, you need to keep updating it to accommodate changing information needs, skill levels and contexts as well as the refinements in project strategies and activities.

The standard criteria for assessing the quality of your M&E system are:

- *Utility* – the M&E system will serve the practical information needs of intended users;
- *Feasibility* – the methods, sequences, timing and processing procedures proposed are realistic, prudent and cost effective;
- *Propriety* – the M&E activities will be conducted legally, ethically and with due regard for the welfare of those affected by its results;
- *Accuracy* – the M&E outputs will reveal and convey technically adequate information.

These criteria can also be used when updating the M&E system.

Assuring quality in M&E

To standardize M&E information in Cuchumatanes, Guatemala, the project – with the implementing agencies – developed formatting outlines and rules about registering and using information. This was critical

need for a process to be able to village and for resources to be They also wanted support to project impacts. In addition, they progress-monitoring formats to qualitative information, besides quantitative information. Finally, they suggested reviewing formats to remove information gathering of data that has not been useful over the last ten years of implementation. Updating not only needs to happen with the project-based M&E systems and procedures. The learning processes of other stakeholder groups also need regular updating.

Just as the project requires continual adaptation, the M&E system will also need to be adjusted regularly and improved as the project evolves and experience develops.

Revising M&E in Guatemala

Four years after start-up, the project team in Cuchumatanes, Guatemala, analyzed both the M&E expectations laid out in the appraisal report and the information needs at different stages of management and

Discussions among key stakeholders are critical to point out weak areas of a project's M&E system. A project in India, discussions with state-level project authorities brought up the document changes from village to available to meet changing priorities. synthesize lessons and document suggested reviewing and modifying provide room for recording

Self-evaluation means that indicators change over time

In one project self-evaluation help groups had become a mundane process and needed to be reviewed. The indicators needed to change, as groups were maturing

Exercise: Participants discuss the possible location of a M&E unit within their health agency. What do they see as the most important logistical issues they will face in setting up and operating a M&E system?

DAY 11 (AM AND PM): EXAMPLE M&E PLANS PREPARED BY PARTICIPANTS

Participants work this day in groups to prepare M&E plan for their health organization.

Trainers serve as resource persons for groups preparing plans. They will be available to help in responding to draft materials prepared by the participants, answering questions about components of the M&E planning process.

DAY 12: (AM) PARTICIPANTS PRESENT M&E PLANS AND DISCUSS

Training participants present M&E plans

Discussion by other participants and trainers

DAY 12 (PM): FINAL THOUGHTS AND SUGGESTED MATERIALS

Final Thoughts: features of a comprehensive national M&E system. Trainers will discuss the various components of the M&E system as a review of the material presented in the training course to reinforce the concepts and the proper application.

Features of a Comprehensive M&E System

M&E UNIT	An established M&E unit within the Ministry of Health
	A budget for M&E that is about 10 percent of the national HIV/AIDS/STI budget
	A significant national contribution to the national M&E budget
	A formalized (M&E) link with the research institutions
	A formalized (M&E) link with leading NGOs and donors
	Epidemiological expertise in the M&E unit or affiliated with the unit
	Behavioral/social science expertise in the M&E unit or affiliated with the unit
	Data processing and statistical expertise in the M&E unit or affiliated with the unit
	Data dissemination expertise in the M&E unit or affiliated with the unit
CLEAR GOALS	Well-defined national program goals and targets
	Regular reviews/evaluations of the progress of the implementation of the national program plans
	Guidelines and guidance to districts and regions or provinces for M&E
	Guidelines for linking M&E to other sectors
	Co-ordination of national and donor M&E needs

INDICATORS	A set of priority indicators and additional indicators at different levels of M&E
	Indicators that are comparable over time
	A number of key indicators that are comparable with other countries
DATA COLLECTION & ANALYSIS	An overall national level data collection and analysis plan
	A plan to collect data and analyze indicators at different levels of M&E
	Second generation surveillance, where behavioral data are linked to HIV/STI surveillance data
DATA DISSEMINATION	An overall national level data dissemination plan
	A well-disseminated informative annual report of the M&E unit
	Annual meetings to disseminate and discuss M&E and research findings with policy-makers and planners
	A clearinghouse for generation and dissemination of findings
	A centralized database or library of all HIV/AIDS/STI-related data collection, including ongoing research
	Co-ordination of national and donor M&E dissemination needs

M&E Sources

Many useful M&E sources can be found on the internet. Below is a listing of useful sources.

American Evaluation Association: www.eval.org

Canadian Evaluation Association: [www. Evaluationcanada.ca](http://www.Evaluationcanada.ca)

Center for Disease Control Evaluation Resources: <http://www.cdc.gov/eval/resources>

European Evaluation Society: <http://europeanevaluation.org/>

Evaluation Center, Western Michigan University: www.wmich.edu/evalctr

Monitoring and Evaluation of Population and Health Programs (MEASURE):

<http://www.cpc.unc.edu.measure/topics/topics.html>

UN Population Fund List of Evaluation Reports and Findings: www.unfpa.org/publications/evaluation

UN Development Project: www.undp.org/eo

World Bank: www.worldbank.org

TRAINING QUESTIONNAIRE

Date:

Name:

Position:

Questions

1. In what ways did you find the training most/least helpful to you in your work?
2. What topics were most useful? Least useful?
3. In what ways do you think you will be able to use this information immediately?
4. In what ways do you think the training will results in changes in the ways in which your health organization is organized?
5. In what ways do you thing the training will change how your health organization will operate, carry out its activities?
6. How do you think the training could have been more useful for you in your work?
7. What specific changes in the training program would you recommend to make it more useful?

Closing remarks:

Thank you for your participation in the training. We hope it has been useful for you in your work. Specifically, we hope that it provided information and ideas that you can use now to improve the design and delivery of health services to the people of Iraq.